

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 13:54:35 ; Search time 120 seconds
(without alignments)
9489.090 Million cell updates/sec

Title: US-10-003-354-3
Perfect score: 3713
Sequence: 1 attaacagcgccgtgtagg.....aaacttaagattatta 3713

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/lna/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/lna/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/lna/PCtUS_COMB.seq:*
6: /cgn2_6/ptodata/2/lna/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.4	1.5	7218	1 US-08-232-463-14	Sequence 14, Appl
2	51.4	1.4	7218	1 US-08-232-463-14	Sequence 14, Appl
3	42.6	1.1	731	1 US-08-451-405A-2	Sequence 2, Appl
4	38.2	1.0	87350	3 US-08-781-891-79	Sequence 79, Appl
5	38.2	1.0	87543	4 US-09-791-211-3	Sequence 3, Appl
6	37.8	1.0	1086	4 US-09-134-001C-1208	Sequence 1208, Ap
7	37.6	1.0	624	4 US-09-397-992A-3	Sequence 3, Appl
8	36	1.0	624	4 US-09-397-992A-6	Sequence 3, Appl
9	36	1.0	1983	4 US-09-435-702B-36	Sequence 36, Appl
10	36	1.0	53526	3 US-08-658-136-2	Sequence 2, Appl
11	36	1.0	53577	3 US-08-658-136-1	Sequence 36, Appl
12	35.8	1.0	549	4 US-09-118-442-26	Sequence 26, Appl
13	35.8	1.0	549	4 US-09-677-064-26	Sequence 26, Appl
14	35.8	1.0	1164	4 US-09-134-001C-2534	Sequence 2534, Ap
15	35.8	1.0	1166	4 US-09-072-596-323	Sequence 323, App
16	35.8	1.0	3095	4 US-09-072-596-323	Sequence 323, App
17	35.8	1.0	19124	2 US-08-487-826B-13	Sequence 13, Appl
18	35.6	1.0	26385	4 US-08-961-527-3	Sequence 3, Appl
19	35.4	1.0	1590	3 US-08-747-221B-23	Sequence 23, Appl
20	35.4	1.0	1590	4 US-09-005-051-23	Sequence 23, Appl
21	35.4	1.0	1650	3 US-08-747-221B-21	Sequence 21, Appl
22	35.4	1.0	1650	3 US-08-747-221B-22	Sequence 21, Appl
23	35.4	1.0	1650	4 US-09-005-051-21	Sequence 22, Appl
24	35.4	1.0	1650	4 US-09-005-051-22	Sequence 22, Appl
25	35.4	1.0	1792	3 US-08-747-221B-18	Sequence 18, Appl
26	35.4	1.0	1792	3 US-08-747-221B-20	Sequence 20, Appl
27	35.4	1.0	1792	4 US-09-005-051-18	Sequence 18, Appl

28	35.4	1.0	1792	4 US-09-005-051-20	Sequence 20, Appl
29	35.4	1.0	2219	5 PCT-US93-07213-4	Sequence 4, Appl
30	35.4	1.0	2870	5 PCT-US93-07213-3	Sequence 3, Appl
31	35.4	1.0	3102	5 PCT-US93-07213-1	Sequence 1, Appl
32	35	0.9	865	4 US-09-040-984-71	Sequence 71, Appl
33	35	0.9	865	4 US-09-123-912-71	Sequence 71, Appl
34	35	0.9	865	4 US-09-643-597-71	Sequence 1, Appl
35	35	0.9	1969	4 US-08-356-340-1	Sequence 1, Appl
36	35	0.9	1969	2 US-08-786-555-1	Sequence 1, Appl
37	35	0.9	2566	2 US-08-488-940-19	Sequence 19, Appl
38	34.6	0.9	35081	2 US-08-752-760A-1	Sequence 1, Appl
39	34.4	0.9	1766	4 US-08-961-527-235	Sequence 235, App
40	34.2	0.9	2659	3 US-08-749-522-1	Sequence 1, Appl
41	34.2	0.9	2755	3 US-08-749-522-2	Sequence 2, Appl
42	34.2	0.9	56516	2 US-08-996-306-1	Sequence 1, Appl
43	34.2	0.9	56516	4 US-09-338-907-1	Sequence 1, Appl
44	34.2	0.9	56516	4 US-09-218-207-1	Sequence 1, Appl
45	34	0.9	421	4 US-08-642-274D-55	Sequence 55, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)856-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14
Query Match 1.5%; Score 55.4; DB 1; Length 7218;

OY	3529
OY	3470
Dd	281
OY	3530
Dd	221
OY	3590
Dd	161
OY	3650
Dd	101
OY	3710
Dd	41

```

07      3449 TTTTTCACAGAGAAGAAATCTTGCGATTGTTTTTTTCATCAATAATTACTATTATGTANG 3508
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
05661 TTATAGTAGTATGAAGAAATTAACAATCAGTTTATTATTCATTAATTCCAAATGCCAAG 45740

```

QY	3509	TATTTAACTGTTTTATTAAAGGCACAAAGTCGCTTACGGGTGGGAGGAATATTTAGGCA	3568
Db	4574.1	GAGAATAATGAAATTAATGGAATGATGTTGTATTACTTGTTCCGATGTGGAAAAAT	4580.0
QY	3569	GGCGCGGCTCTAGGCAAAAGCATGGGAGCAGCATTTTTTAAAGTTCATCTTTGCG	3628
Db	4580.1	TTATATGCTCTTTCTCAAACAAGCACACTGTCCAGTAGAAAATCAATGTGAGCTACATATGCCA	4586.0
QY	3629	CTTACTTTGTAATYTGTCCAGAAATGGCAATACAAATATPAAAAAGTATATATGTTTTAAT	3688
Db	4586.1	TTTTTAAATTTCTACATAGCCACATTTTAAAAAGTAAAGATGACAAATTAATTTTGATAAT	4592.0
QY	3689	GTAATAAACCT	3699
Db	4592.1	ATAATTAAAT	4593.1

```

1 RESULT 5
2 US-09-791-211-3
3 Sequence 3, Application US/09791211
4 Patent No. 6448080
5 GENERAL INFORMATION:
6 APPLICANT: Donna T. Ward
7 APPLICANT: Andrew T. Watt
8 TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
9 FILE REFERENCE: RTS-0205
10 CURRENT APPLICATION NUMBER: US/09/791,211
11 NUMBER OF SEQ ID NOS: 2001-02-23
12 SEQ ID NO 3
13 LENGTH: 87543
14 TYPE: DNA
15 ORGANISM: Homo sapiens
16 FEATURE:
17 NAME/KEY: unsure
18 LOCATION: 7421
19 OTHER INFORMATION: unknown
20 NAME/KEY: unsure
21 LOCATION: 7427
22 OTHER INFORMATION: unknown
23 NAME/KEY: unsure
24 LOCATION: 11609
25 OTHER INFORMATION: unknown
26 NAME/KEY: unsure
27 LOCATION: 12605
28 OTHER INFORMATION: unknown
29 NAME/KEY: unsure
30 LOCATION: 12742
31 OTHER INFORMATION: unknown
32 NAME/KEY: unsure
33 LOCATION: 29370
34 OTHER INFORMATION: unknown
35 NAME/KEY: unsure
36 LOCATION: 29422
37 OTHER INFORMATION: unknown
38 NAME/KEY: unsure
39 LOCATION: 29979
40 OTHER INFORMATION: unknown
41 NAME/KEY: unsure
42 LOCATION: 29980
43 OTHER INFORMATION: unknown
44 NAME/KEY: unsure
45 LOCATION: 29981
46 OTHER INFORMATION: unknown
47 NAME/KEY: unsure
48 LOCATION: 30136
49 OTHER INFORMATION: unknown
50 NAME/KEY: unsure
51 LOCATION: 30140
52 OTHER INFORMATION: unknown
53 NAME/KEY: unsure
54 LOCATION: 31205
55

```



```
Query Match 1.0%; Score 37.8; DB 4; Length 1086;
Best Local Similarity 55.8%; Pred. No. 0.51;
Matches 72; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

OY 2916 TTTTGCAATTTGAAATATATCTGCTTTTCTTAATGAGACACTTACCAATGAT 2975
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 385 TATTCATATCATATATATGACATTCCTTCTTATCATCATTTACAAATTAAT 444
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2976 TTTGACATCTCCAGAGGATTTTTCCTCTCCATCTTCCACAGCTCT 3035
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 445 TTTGATTTGTTCAATAGTAAATTTGCTGCTTTAGCATTTAAGTCAACAAATCTTAT 504
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3036 CCTGTTGT 3044
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 505 CTACATTAAT 513
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-09-397-992A-3/c
: Sequence 3, Application US/09397992A
: Patent No. 6329175
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell
: APPLICANT: Grant, Francis J.
: APPLICANT: Rixon, Mark W.
: APPLICANT: Kindsvogel, Wayne
: TITLE OF INVENTION: Interferon-epsilon
: FILE REFERENCE: 98-46
: CURRENT APPLICATION NUMBER: US/09/397,992A
: CURRENT FILING DATE: 1999-09-16
: PRIOR APPLICATION NUMBER: 60/101,012
: PRIOR FILING DATE: 1998-09-18
: PRIOR APPLICATION NUMBER: 60/118,578
: PRIOR FILING DATE: 1999-02-05
: PRIOR APPLICATION NUMBER: 60/142,766
: PRIOR FILING DATE: 1999-07-08
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 624
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: This degenerate sequence encodes the amino acid
: OTHER INFORMATION: sequence of SEQ ID NO:2.
: NAME/KEY: variation
: LOCATION: (1)...(624)
: OTHER INFORMATION: n is any nucleotide
US-09-397-992A-3

Query Match 1.0%; Score 37.6; DB 4; Length 624;
Best Local Similarity 27.2%; Pred. No. 0.4;
Matches 72; Conservative 60; Mismatches 130; Indels 3; Gaps 1;

OY 3421 TTTCTGGCTGCTCTTGTGTCATGTTTTCACAGAGGAAGAAATTTGCTATT 3480
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 310 TTTCTGCCANCCRCNARNRNSWDATRTTNGCNCKRAANRNSWRADATYTTGNARCA 251
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3481 TTTTTCATATTAATGATTTATGATGATTTAAGTGTATTATTAAGACAGATTCG 3540
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 250 TTTCTGNAARATNGCNARNRNGTCCTTTT--GRTATYGYNGNGNSNARNRNSWT 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3541 TTAGGGGTGGAGGAATATTTGAGGAGGCTGCTTTAGGAAGAAGATGGGAAGC 3600
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 193 TTTGNGNARNRARNRARTTTTNCRTGNGNARCAATGTGDTNSNARNNGYTTGNA 134
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3601 AACATTTTATTAAGTGTACTATTGCTCTACTTGTATGTTGTCAGAAATGCCAATA 3660
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 RYTTTNTNARNRYTTNARNRNSWYTCYGTGRTTACYTGCKYGYGGRADATDATNARYT 74
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3661 CAATATAAAGCATATATGTTT 3685
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB 73 TNARRTCNARNRNSWRADATNGTNGT 49

RESULT 8
US-09-397-992A-6/c
: Sequence 6, Application US/09397992A
: Patent No. 6329175
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell
: APPLICANT: Grant, Francis J.
: APPLICANT: Rixon, Mark W.
: APPLICANT: Kindsvogel, Wayne
: TITLE OF INVENTION: Interferon-epsilon
: FILE REFERENCE: 98-46
: CURRENT APPLICATION NUMBER: US/09/397,992A
: CURRENT FILING DATE: 1999-09-16
: PRIOR APPLICATION NUMBER: 60/101,012
: PRIOR FILING DATE: 1998-09-18
: PRIOR APPLICATION NUMBER: 60/118,578
: PRIOR FILING DATE: 1999-02-05
: PRIOR APPLICATION NUMBER: 60/142,766
: PRIOR FILING DATE: 1999-07-08
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 6
: LENGTH: 624
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: This degenerate sequence encodes the amino acid
: OTHER INFORMATION: sequence of SEQ ID NO:5.
: NAME/KEY: variation
: LOCATION: (1)...(624)
: OTHER INFORMATION: n is any nucleotide
US-09-397-992A-6

Query Match 1.0%; Score 36; DB 4; Length 624;
Best Local Similarity 26.8%; Pred. No. 1.2;
Matches 71; Conservative 60; Mismatches 131; Indels 3; Gaps 1;

OY 3421 TTTCTGGCTGCTCTTGTGTCATGTTTTCACAGAGGAAGAAATTTGCTATT 3480
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 310 TTTCTGCCANCCRCNARNRNSWDATRTTNGCNCKRAANRNSWRADATYTTGNARCA 251
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3481 TTTTTCATATTAATGATTTATGATGATTTAAGTGTATTATTAAGACAGATTCG 3540
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 250 TTTCTGNAARATNGCNARNRNGTCCTTTT--GRTATYGYNGNGNSNARNRNSWT 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3541 TTAGGGGTGGAGGAATATTTGAGGAGGCTGCTTTAGGAAGAAGATGGGAAGC 3600
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 193 TTTGNGNARNRARNRARTTTTNCRTGNGNARCAATGTGDTNSNARNNGYTTGNA 134
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3601 AACATTTTATTAAGTGTACTATTGCTCTACTTGTATGTTGTCAGAAATGCCAATA 3660
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 RYTTTNTNARNRYTTNARNRNSWYTCYGTGRTTACYTGCKYGYGGRADATDATNARYT 74
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 3661 CAATATAAAGCATATATGTTT 3685
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 73 TNARRTCNARNRNSWRADATNGTNGT 49

RESULT 9
US-09-453-702B-36
: Sequence 36, Application US/09453702B
: Patent No. 6365723
: GENERAL INFORMATION:
: APPLICANT: Blattner, Frederick R.
: APPLICANT: Burland, Valerie
: APPLICANT: Perna, Nicole T.
: APPLICANT: Plunkett, Guy
: APPLICANT: Welch, Rod
: TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
: NUMBER OF SEQUENCES: 265
```


ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GENA-17, 8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 53577 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-658-136-1

Query Match 1.0%; Score 36; DB 3; Length 53577;
Best Local Similarity 51.2%; Pred. No. 26;
Matches 84; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 241 AGAGCCCGCTCGACGTCTGTGAGGAGCCCGGAGGGGGGAGGTGCGCCACAGAA 300
DB 35339 AGAGGGGGATGGATGGATGAGAGGAGAGGGGAGGAGGGGATGGAGAGGGGAGGA 35280
QY 301 GCGGGCTCTGTAAAGACAGCTTGGGAAGATTCCGAGAGAGAGGAAGACCGGATT 360
DB 35279 GCGGGCAAGGAG 35220
QY 361 GAAAGAGAGCCAGCGCTGAGGGGAGGGGCGCTCTAAGATTG 404
DB 35219 GGAGAGAGAGAGGGGGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 35176

RESULT 12
US-09-118-442-26
Sequence 26, Application US/09118442B
Patent No. 6197561
GENERAL INFORMATION:
APPLICANT: Martino-Catt, Susan J.
APPLICANT: Wang, Hongyu
APPLICANT: Beach, Larry R.
APPLICANT: Wang, Xun
TITLE OF INVENTION: Genes Controlling Phylate Metabolism In
TITLE OF INVENTION: Plants and Uses Thereof
FILE REFERENCE: 0706
CURRENT APPLICATION NUMBER: US/09/118,442B
EARLIER FILING DATE: 1998-07-17
EARLIER APPLICATION NUMBER: 60/055,446
EARLIER FILING DATE: 1997-08-11
EARLIER APPLICATION NUMBER: 60/055,526
EARLIER FILING DATE: 1997-08-08
EARLIER APPLICATION NUMBER: 60/053,944
EARLIER FILING DATE: 1997-07-28
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 26
LENGTH: 549
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(549)
OTHER INFORMATION: n = A,T,C or G
US-09-118-442-26

Query Match 1.0%; Score 35.8; DB 4; Length 549;
Best Local Similarity 53.1%; Pred. No. 1.3;
Matches 76; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1551 TGCTTATATGGCATCTTGCATCTTACAGCTTGTGTAAGAGTTGGAGC 1610
DB 213 TCCTATCTTTGGATATCGACATCTCCAGATTACGTTAAGAAAGCGGGCGAGC 272

QY 1611 ACTCTGGAAACCCCTGTGATGACGAGACATGCTCTCACTGCATGCCAGGCTTCT 1670
DB 273 ATGCTTACAAAGTCATTACAGACATCCCACTGATCTCTCCGCGAGACCGAGCTCT 332
QY 1671 ACGCTGAAGGTTCCAGCGCTTC 1693
DB 333 ACTGGAAGATTCCAGACTTC 355

RESULT 13
US-09-677-064-26
Sequence 26, Application US/09677064
Patent No. 6291224
GENERAL INFORMATION:
APPLICANT: Martino-Catt, Susan J.
APPLICANT: Wang, Hongyu
APPLICANT: Beach, Larry R.
TITLE OF INVENTION: Genes Controlling Phylate Metabolism In
TITLE OF INVENTION: Plants and Uses Thereof
FILE REFERENCE: 0706D
CURRENT APPLICATION NUMBER: US/09/677,064
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/055,446
PRIOR FILING DATE: 1997-08-11
PRIOR APPLICATION NUMBER: 60/055,526
PRIOR FILING DATE: 1997-08-08
PRIOR APPLICATION NUMBER: 60/053,944
PRIOR FILING DATE: 1997-07-28
PRIOR APPLICATION NUMBER: 09/118,442
PRIOR FILING DATE: 1998-07-17
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 26
LENGTH: 549
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(549)
OTHER INFORMATION: n = A,T,C or G
US-09-677-064-26

Query Match 1.0%; Score 35.8; DB 4; Length 549;
Best Local Similarity 53.1%; Pred. No. 1.3;
Matches 76; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1551 TGCTTATATGGCATCTTGCATCTTACAGCTTGTGTAAGAGTTGGAGC 1610
DB 213 TCCTATCTTTGGATATCGACATCTCCAGATTACGTTAAGAAAGCGGGCGAGC 272
QY 1611 ACTCTGGAAACCCCTGTGATGACGAGACATGCTCTCACTGCATGCCAGGCTTCT 1670
DB 273 ATGCTTACAAAGTCATTACAGACATCCCACTGATCTCTCCGCGAGACCGAGCTTC 332
QY 1671 ACGCTGAAGGTTCCAGCGCTTC 1693
DB 333 ACTGGAAGATTCCAGACTTC 355

RESULT 14
US-09-134-001C-2534/C
Sequence 2534, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 2534
LENGTH: 1164
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2534

Query Match 1.0%; Score 35.8; DB 4; Length 1164;
Best Local Similarity 49.7%; Pred. No. 2.1;
Matches 91; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 957 AAGAGCGGAAATTTCTGCAAGATGCTTCCAGAGATACATGACCAAGCAAGACC 1016
DB 759 AACATAGAAATGCTGACCAATATTTCTTCAACACTTTCTTCTGACTGCTTC 700
QY 1017 CTCGACTTTGCTCCCTAAATTCATGACTGTGCTGACGAGGAGGCAAGACA 1076
DB 699 TTCTATTAAGAGGTTTAAAGTCTGCCAAGGTTCTGAGAGTAAGTTGAGTTTAAATAC 640
QY 1077 TTCCGATTTGCTGATTAACAATCTTTTACCAAGATGCGTAAATATCAATATG 1136
DB 639 TTCTATTTGATTAATAATTTTCACTTCTGAAAGGTAATAATTTCTTGGGAATATT 580
QY 1137 ACC 1139
DB 579 ACC 577

RESULT 15
US-09-072-596-323/C
Sequence 323, Application US/09072596
Patent No. 6458366

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESS: SPED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 323:
SEQUENCE CHARACTERISTICS:
LENGTH: 1166 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-072-596-323

Query Match 1.0%; Score 35.8; DB 4; Length 1166;
Best Local Similarity 29.1%; Pred. No. 2.1;
Matches 83; Conservative 54; Mismatches 148; Indels 0; Gaps 0;

QY 163 GCGGTTAACTGTGAGAGGGGTGGGGACGTGATTTCTTCCCATGGCAGGCAATG 222
DB 559 GAGGGGAATNMGTSGKRGAGADGMTSGKRAATNGCTGCMAGNGTGAAGTGMSSG 500
QY 223 GTGGCTTGAGCTGTGTCAGAGCCGCTCGACTGTCTGAGGAGGCCCGAGGGGCG 282
DB 499 GKKANAGGWAACGGGANKKAFRTGRKXKMGAGGGRCAAGTGKMTGAGRCSG 440
QY 283 GGGAGGTGCCACACAGAACCGCGTTCTTAAGAGACCTTGGAAGATGATTCGAG 342
DB 439 KGRGRATGACGAGGGRTGWSMGRTWCGGGGKRNRAKTAATCCNNGRAKAGASACRMK 380
QY 343 AAGAGGAAGAACCGGATTTGAAGAAGAGCCGCTGAGGGGAGGGGCTGTAAGAT 402
DB 379 SAKTTSMGKSSAGNKKAAAGAMCNAARWNNCCGTTSCCGRGTRRRSAAACSAAT 320
QY 403 GCGGTGCGCTCCTCCGGCCGTCGTTTGGTGGGTTTATC 447
DB 319 CCCGTTACCGGTATGTAACCAACCGGACCGGTCGARTTATAC 275

Search completed: January 8, 2003, 18:17:00
Job time : 632 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 8, 2003, 12:53:14 ; Search time 716 Seconds
(without alignments)
11678.310 Million cell updates/sec

Title: US-10-003-354-3

Perfect score: 3713
Sequence: 1 attacagcgcgcgtgtagg.....aaacttaatgattatta 3713

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3702	99.7	3714	23 AAS73986
2	2014	54.2	2133	23 AAS73982
3	1852	49.9	2010	23 AAS73983
4	1325.2	35.7	1992	23 AAS73985
5	581.6	15.7	2764	18 AAT84358
6	469.8	12.7	598	22 ABA62045
7	469.8	12.7	598	22 AAK10359
8	469.8	12.7	598	22 AAK36262
9	469.8	12.7	598	22 AAI41984

10	408	11.0	1291	22 AAB33511	Human colon cancer
11	408	11.0	1899	24 ABL65517	Lung cancer relate
12	357.4	9.6	1795	23 ABL06531	Drosophila melanog
13	346.6	9.3	3309	23 ABL15963	Drosophila melanog
14	346.6	9.3	6699	23 ABL15962	Drosophila melanog
15	324.8	8.7	674	22 AAB73319	Human cervical can
16	324.8	8.7	674	22 AAB73320	Human cervical can
17	303.6	8.2	399	21 AAC04404	Human foetal liver
18	285	7.7	346	22 ABA74539	Human foetal liver
19	285	7.7	346	22 AAK23006	Human foetal liver
20	285	7.7	346	22 AAK49183	Human bone marrow
21	285	7.7	346	22 AAI55018	Probe #23704 used
22	250.6	6.7	885	23 AAS68858	DNA encoding novel
23	179	4.8	4687	23 ABL06530	Drosophila melanog
24	164.8	4.4	300	18 AAT84371	Friedreich's ataxi
25	159	4.3	449	23 AAS68856	DNA encoding novel
26	159	4.3	449	23 AAS73984	DNA encoding novel
27	101.6	2.7	564	22 AAB34797	Human colon cancer
28	98.6	2.7	153	18 AAT84370	Friedreich's ataxi
29	91.6	2.5	3743	24 ABL69117	Kidney cancer rela
30	85.8	2.3	1813	21 AAC39125	Arabidopsis thalia
31	79.8	2.1	1099	21 AAC41576	Arabidopsis thalia
32	79.8	2.1	1707	21 AAC41671	Arabidopsis thalia
33	77	2.1	118	18 AAT84369	Friedreich's ataxi
34	75	2.0	2481	21 AAC48385	Arabidopsis thalia
35	75	2.0	2580	21 AAC42792	Arabidopsis thalia
36	72.2	1.9	3176	22 AAB42605	Nucleotide sequenc
37	72	1.9	1219	23 ABL18985	Drosophila melanog
38	70.2	1.9	131	18 AAT84368	Friedreich's ataxi
39	65.8	1.8	174	24 AAB26423	Human ORFX polynuc
40	63.4	1.7	2872	22 AAF93889	Human cDNA encodin
41	61.4	1.7	211	18 AAT84372	Friedreich's ataxi
42	60	1.6	60	24 ABA41013	Human spliced tran
43	58	1.5	183	21 AAC09007	Human secreted pro
44	50	1.3	537	23 ABA45606	Human prostate exp
45	50	1.3	2095	23 ABA30111	Human prostate exp

ALIGNMENTS

RESULT 1	
AAS73986	AAS73986 standard; cDNA; 3714 BP.
ID	
AC	AAS73986;
DT	13-FEB-2002 (first entry)
DE	DNA encoding novel human diagnostic protein #9790.
DE	Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
KW	Homo sapiens.
OS	WO200175067-A2.
PN	11-OCT-2001.
PD	30-MAR-2001; 2001WO-US08631.
PF	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	(HYSE-) HYSEQ INC.
XX	Dymanac RT, Liu C, Tang YT;
XX	WPI: 2001-639362/73.
XX	P-PSDB; ABB09799.
XX	New isolated polynucleotide and encoded polypeptides, useful in

QY 1680 GGTTCACACCGCTTCATGTGCAACACAGTATTTAAGACATTCCTTTGAACCTTCTCCTT 1739
| | | | |
Db 1681 GGTTCACACCGCTTCATGTGCAACACAGTATTTAAGACATTCCTTTGAACCTTCTCCTT 1740
| | | | |
QY 1740 CCAAAAATTTGCGTGTGCTCATCTTCTCTCGGCGAGCAGGCTCCAGTGGCACTCCT 1799
| | | | |
Db 1741 CCAAAAATTTGCGTGTGCTCATCTTCTCTCGGCGAGCAGGCTCCAGTGGCACTCCT 1800
| | | | |
QY 1800 GCATTACTTACACAGCATCGGTCTCTGGGGAACACACAGGCAACAGTGCACAAAGGAG 1859
| | | | |
Db 1801 GCATTACTTACACAGCATCGGTCTCTGGGGAACACACAGGCAACAGTGCACAAAGGAG 1860
| | | | |
QY 1860 AAGTGGACCCAGGCGCTTACCTTGCTCTCGATGATTTTACCTCAGACTCCACTTGG 1919
| | | | |
Db 1861 AAGTGGACCCAGGCGCTTACCTTGCTCTCGATGATTTTACCTCAGACTCCACTTGG 1920
| | | | |
QY 1920 AGGAATAGTGAAGGCGCTATCTGACCCAGTTTCACACTCAGTGTGAGAGA 1979
| | | | |
Db 1921 AGGAATAGTGAAGGCGCTATCTGACCCAGTTTCACACTCAGTGTGAGAGA 1980
| | | | |
QY 1980 CTTTGCAATGCTAATCAAGTACACACCTTGGAAACCTTGAAGTGCAGATGAGAT 2039
| | | | |
Db 1981 CTTTGCAATGCTAATCAAGTACACACCTTGGAAACCTTGAAGTGCAGATGAGAT 2040
| | | | |
QY 2040 TCACCCATTAAGCGCAAGCCTCAGACACCTGGAACAGATTCGCATCTGTGATC 2099
| | | | |
Db 2041 TCACCCATTAAGCGCAAGCCTCAGACACCTGGAACAGATTCGCATCTGTGATC 2100
| | | | |
QY 2100 CCAAGATTCAGCCCTTGCCCGACAGATTCGAATTTCTTCTACTTGGTCAATCAAAAA 2159
| | | | |
Db 2101 CCAAGATTCAGCCCTTGCCCGACAGATTCGAATTTCTTCTACTTGGTCAATCAAAAA 2160
| | | | |
QY 2160 GGAGTGTATAGAAAGTGAAGGAGCTGCTCCATCTTCTTCTCTGAAAGAAACCTTCT 2219
| | | | |
Db 2161 GGAGTGTATAGAAAGTGAAGGAGCTGCTCCATCTTCTTCTCTGAAAGAAACCTTCT 2220
| | | | |
QY 2220 CTCTCTCTCTCTCTCATGAAATGGCGCTTGAAGCTCAGAGATGAGAGACCGACATC 2279
| | | | |
Db 2221 CTCTCTCTCTCTCTCATGAAATGGCGCTTGAAGCTCAGAGATGAGAGACCGACATC 2280
| | | | |
QY 2280 CCCTCCATCCAGAGTGGGAGTGGATTTTCAACTGGCCCAACCTTGGCTCCACTA 2339
| | | | |
Db 2281 CCCTCCATCCAGAGTGGGAGTGGATTTTCAACTGGCCCAACCTTGGCTCCACTA 2340
| | | | |
QY 2340 TTGAATTTTTCAGACCCCACTTCTCATGCTGGAATGGATTTGCTGACTTGGACAC 2399
| | | | |
Db 2341 TTGAATTTTTCAGACCCCACTTCTCATGCTGGAATGGATTTGCTGACTTGGACAC 2400
| | | | |
QY 2400 TTTCTTTCCCTCGTCTTTCAGTGAACCGGACTCTTAATTTCTCTCAGACAGACTAGC 2459
| | | | |
Db 2401 TTTCTTTCCCTCGTCTTTCAGTGAACCGGACTCTTAATTTCTCTCAGACAGACTAGC 2460
| | | | |
QY 2460 TGGCAATTAATCCCTACTTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2519
| | | | |
Db 2461 TGGCAATTAATCCCTACTTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520
| | | | |
QY 2520 CTCTGTAAGGTTTTCGGGGTAAGGTTTAAACCACTCCAGCTTCTCTCTCTCT 2579
| | | | |
Db 2521 CTCTGTAAGGTTTTCGGGGTAAGGTTTAAACCACTCCAGCTTCTCTCTCTCT 2580
| | | | |
QY 2580 TTTTCTTTCTGAAAAAAGGAAAGGACACACAGACAAATTTTCACATCTCAG 2639
| | | | |
Db 2581 TTTTCTTTCTGAAAAAAGGAAAGGACACACAGACAAATTTTCACATCTCAG 2640
| | | | |
QY 2640 AACCTCAGAAAGTGTGAAAGATGCTATTTGTAAGTTCCCTGGAAGGCAAGGTGT 2699
| | | | |
Db 2641 AACCTCAGAAAGTGTGAAAGATGCTATTTGTAAGTTCCCTGGAAGGCAAGGTGT 2700
| | | | |
QY 2700 TTATGAGAGAGAGTAGTGAATGCTCTGCGAAGACAGCTCTTTTAAATCCTCTC 2759
| | | | |
Db 2701 TTATGAGAGAGAGTAGTGAATGCTCTGCGAAGACAGCTCTTTTAAATCCTCTC 2760
| | | | |
QY 2760 TCTTGATGAATTTCTTAAGGCTGAAGAAATGAAGAGACTGGGACATGGGCTAATCTTAT 2819
| | | | |

Db 2761 TCTTGATGAATTTCTTAAGGCTGAAGAAATGAAGAGAGGACATGGGCTAATCTTAT 2820
| | | | |
QY 2820 CCCTTTTGTAAACAGAGGAGGACCATATGGCTGGGAGATCATAGCCCTCTTAGCAGA 2879
| | | | |
Db 2821 CCCTTTTGTAAACAGAGGAGGACCATATGGCTGGGAGATCATAGCCCTCTTAGCAGA 2880
| | | | |
QY 2880 ATCTGTTCACTGCCAGGCTATAGTAATTAATTAATTTTGAATTTGAATTAATTCG 2939
| | | | |
Db 2881 ATCTGTTCACTGCCAGGCTATAGTAATTAATTAATTTTGAATTTGAATTAATTCG 2940
| | | | |
QY 2940 GTTGTTTTCTTAATGTGAAGACTTACCAATGAATTTTGAATTCCTCAGAGAGAT 2999
| | | | |
Db 2941 GTTGTTTTCTTAATGTGAAGACTTACCAATGAATTTTGAATTCCTCAGAGAGAT 3000
| | | | |
QY 3000 TTTTGTGCTCTCTCATCTTTTCCAAAGGCTCTCTCTGTTGGAGCTAAGGTAAAG 3059
| | | | |
Db 3001 TTTTGTGCTCTCTCATCTTTTCCAAAGGCTCTCTCTGTTGGAGCTAAGGTAAAG 3060
| | | | |
QY 3060 AGGGACACTTCTGCTGTTTAAACAGACAGTCCATATCTGAGGCCAGCAATATTTTC 3119
| | | | |
Db 3061 AGGGACACTTCTGCTGTTTAAACAGACAGTCCATATCTGAGGCCAGCAATATTTTC 3120
| | | | |
QY 3120 TTAACATCATGGGAGACAGCAGATTCCTTGGTGAGTCAATTCCTGACATATGT 3179
| | | | |
Db 3121 TTAACATCATGGGAGACAGCAGATTCCTTGGTGAGTCAATTCCTGACATATGT 3180
| | | | |
QY 3180 CCTACCCCTCTCTCTCATGAGGAGGAAATGGGAGGCTACATATGCTCTCTC 3239
| | | | |
Db 3181 CCTACCCCTCTCTCTCATGAGGAGGAAATGGGAGGCTACATATGCTCTCTC 3240
| | | | |
QY 3240 CCCGCTACAAAGATTTGTTTCCATCTGATCCCTTCTCTGAGGGAAGAG 3299
| | | | |
Db 3241 CCCGCTACAAAGATTTGTTTCCATCTGATCCCTTCTCTGAGGGAAGAG 3300
| | | | |
QY 3300 GGGCTGCTATCTCAGGCAAGATTTGAAATTCCTGTTCTATCCCTTCTATCCACCT 3359
| | | | |
Db 3301 GGGCTGCTATCTCAGGCAAGATTTGAAATTCCTGTTCTATCCCTTCTATCCACCT 3360
| | | | |
QY 3360 GCTTGATTAATATGTTAGGCCATACCCCAATTAATGCTATTAAGACCCCAAGCA 3419
| | | | |
Db 3361 GCTTGATTAATATGTTAGGCCATACCCCAATTAATGCTATTAAGACCCCAAGCA 3420
| | | | |
QY 3420 GTTCTGCTGCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3479
| | | | |
Db 3421 GTTCTGCTGCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3480
| | | | |
QY 3480 TTTTCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3539
| | | | |
Db 3481 TTTTCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3540
| | | | |
QY 3540 GTTAGGGGTGGAGGGAATTTAGAGGAGGCTGGGCTTTAGGGAAGAAATGGGAAG 3599
| | | | |
Db 3541 GTTAGGGGTGGAGGGAATTTAGAGGAGGCTGGGCTTTAGGGAAGAAATGGGAAG 3600
| | | | |
QY 3600 CAACATTTTATTAAGTGTACTATTTGCTACTTGTATTTGCTCAGAAATGGCAAT 3659
| | | | |
Db 3601 CAACATTTTATTAAGTGTACTATTTGCTACTTGTATTTGCTCAGAAATGGCAAT 3660
| | | | |
QY 3660 ACAATTAATAAGTGAATATGTTTAAATGAATAAATTTAATGAATTAATGAATTAAT 3713
| | | | |
Db 3661 ACAATTAATAAGTGAATATGTTTAAATGAATAAATTTAATGAATTAATGAATTAAT 3714
| | | | |
RESULT 2
AAS73982
ID AAS73982 standard; cDNA, 2133 BP.
XX
AC AAS73982;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #9786.

XX	Human: chromosome mapping; gene mapping; gene therapy; forensic:
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
XX	30-MAR-2001; 2001WO-US08631.
PF	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drimaenac RT, Liu C, Tang YT;
XX	
DR	WPI: 2001-639362/73.
XX	P-PSDB; ABG09795.
DR	
XX	
XX	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
PS	Claim 1: SEQ ID No 9786; 103pp: English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS64197-AAS94564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
XX	
SQL	Sequence 2133 BP; 524 A; 524 C; 559 G; 526 T; 0 other;
XX	
Query Match	54.2%; Score 2014; DB 23; Length 2133;
Best Local Similarity	98.2%; Pred. No. 0;
Matches 2073; Conservative	0; Mismatches 0; Indels 39; Gaps 2;
QY	1 ATTACAGGCCCTGGTTAGAAGAGACGGAGGAGCGCTCGCTCTTGGACCTTTCA 60
DB	
QY	22 ATTAAAGGCGCTGGTTAGGAAGACGGAGGAGCGCTCGCTCTTGGACCTTTCA 81
QY	61 TGCCTCGTTTTTTTTCACATGTCGAGCTTGCTGGGCGCAAGGTCCACAGCAGCACTTA 120
DB	
QY	82 TGCCTCGTTTTTTTTCACATGTCGAGCTTGCTGGGCGCAAGGTCCACAGCAGCACTTA 141
QY	121 AGCTTACTCTTGTGTAAGGGGAAAGTATCCCTCTGTGGAAGCGGTAACTTGTGAG 180
DB	
QY	142 AGCTTACTCTTGTGTAAGGGGAAAGTATCCCTCTGTGGAAGCGGTAACTTGTGAG 201
QY	181 GGGGTCGGGAGAGTGTGTTCCCATGCCAGAGCAATGGTGTGGCCCTTGAAGTGTGC 240
DB	
QY	202 GGGGTCGGGAGAGTGTGTTCCCATGCCAGAGCAATGGTGTGGCCCTTGAAGTGTGC 261

OY	241	AGGAGCCGGGCTCGACGTGTGTGAGAGAGGCGCCGAGAGGGGCGGGAGGGTGGCCACAA	300
Db	262	AGGAGCCGGGCTCGACGTGTGTGAGAGAGGCGCCGAGAGGGGCGGGAGGGTGGCCACAA	321
OY	301	CGCGGGTCTGTAAAGACGTTGGGAGATTGATTCGAGAGAGAGAACCGGATT	360
Db	322	CGCGGGTCTGTAAAGACGTTGGGAGATTGATTCGAGAGAGAGAACCGGATT	381
OY	361	GAAGAAGGCCAGGCGCTAGAGGGGGGGGGGGTGTAGATAGGGGTGCGCTCCTCGG	420
Db	382	GAAGAAGGCCAGGCGCTAGAGGGGGGGGGGGTGTAGATAGGGGTGCGCTCCTCGG	441
OY	421	GCCGTCGTCTCGGTCGGTTTTTTCATCTTGAATCCCGGGTCCCTCCTGACCTTTC	480
Db	442	GCCGTCGTCTCGGTCGGTTTTTTCATCTTGAATCCCGGGTCCCTCCTGACCTTTC	501
OY	481	CT-----CAGCATCTGGATCAAGAGACCCATGCGATCTGA-----	516
Db	502	CTCAGCAGCATCTGGATCAAGAGACCCATGCGATCTGAGGCTTGAAGCTAGACAGGA	561
OY	517	-----GGTGCCTATGCGCTGTGCGATCGCCATCAAGAAATAGGCGCATAG	561
Db	562	TTCTTACATCTCATTTGGTGGCTTATGCTCTGTGCGATGCGCCATCAAGAAATAGGCGCATAG	621
OY	562	AAGTGTGATTCCTCAGAGAGAGACAACATATAAAAAGACAACTCATGACGCTTGAAGG	621
Db	622	AAGTGTGATTCCTCAGAGAGAGACAACATATAAAAAGACAACTCATGACGCTTGAAGG	681
OY	622	TGCCATCCAGTTAGGCAATTCACACCTGTGGGGACCGTGAATACCAAAACGAGCGTGA	681
Db	682	TGCCATCCAGTTAGGCAATTCACACCTGTGGGGACCGTGAATACCAAAACGAGCGTGA	741
OY	682	TGTCTCATGCAAGATTCTACGTGGTGTGAGATCTTCTTCCAGTGAAGGGAGCA	741
Db	742	TGTCTCATGCAAGATTCTACGTGGTGTGAGATCTTCTTCCAGTGAAGGGAGCA	801
OY	742	CCTGACCCCTGCATACATCAATGACTTGGTTCAAGACTTATGACCGTGGCTT	801
Db	802	CCTGACCCCTGCATACATCAATGACTTGGTTCAAGACTTATGACCGTGGCTT	861
OY	802	CCGCTACTTCCGGAGGCTATTTGGTATCCGCGCCGAGTATCTGTATCCCTCGCAG	861
Db	862	CCGCTACTTCCGGAGGCTATTTGGTATCCGCGCCGAGTATCTGTATCCCTCGCAG	921
OY	862	TGAGCGCGCTGATTAAGCTCTGTAGCTGTGAGCTAGTGGTCCCTATCTATGTCCAG	921
Db	922	TGAGCGCGCTGATTAAGCTCTGTAGCTGTGAGCTAGTGGTCCCTATCTATGTCCAG	981
OY	922	CGACGATAGTTCAATTTAAAGACAGTCCAACTTAAAGAGGGGGAATTTCTGCAACA	981
Db	982	CGACGATAGTTCAATTTAAAGACAGTCCAACTTAAAGAGGGGGAATTTCTGCAACA	1041
OY	982	GCTTCAGGATCTACATGAAGCTCAACAGAAACCTCGGACTTGTGCGCTAAATTTCTA	1041
Db	1042	GCTTCAGGATCTACATGAAGCTCAACAGAAACCTCGGACTTGTGCGCTAAATTTCTA	1101
OY	1042	TGGAATGTACGTGTGCAAGAGAGGTGGCAAGAACATTCGATTTGGTGTATATACAACT	1101
Db	1102	TGGAATGTACGTGTGCAAGAGAGGTGGCAAGAACATTCGATTTGGTGTATATACAACT	1161
OY	1102	TTTACCAAGATCGGTAAAAAATGCAATTCAAATATGACCTCAAAAGGCTCAAGCTA	1161
Db	1162	TTTACCAAGATCGGTAAAAAATGCAATTCAAATATGACCTCAAAAGGCTCAAGCTA	1221
OY	1162	GGGGCTTCCAGAAAGAGGAGAGAGCTTCCACATTTAAAGACTTGAAGCTTCTT	1221
Db	1222	GGGGCTTCCAGAAAGAGGAGAGAGCTTCCACATTTAAAGACTTGAAGCTTCTT	1281
OY	1222	ACAACAATCCCTGATGGTCTTTTGTGATGCGATATGACAAAGCTCTGTGTAAC	1281
Db	1282	ACAACAATCCCTGATGGTCTTTTGTGATGCGATATGACAAAGCTCTGTGTAAC	1341
OY	1282	CGTGAGCGTACGTGTTTGGTGTGCGAGACTTCAAGATATATGATTTACAGCTCTTGAT	1341

Oy	361	GAAGAGAGCGCAGGCCGCTGAGAGGGGAGGGGGGCTGTAAGATGGCGTCGCGGCTCCGCGG	420
Db	362	GAAGAGAGCGCAGGCCGCTGAGAGGGGAGGGGGGCTGTAAGATGGCGTCGCGGCTCCGCGG	421
Oy	421	GGCGTCGCTCGGTCGCTGCTTTTTCATCCCTTGATCCCGCGGCTCCCTGCTGACTGTGC	480
Db	422	GGCGTCGCTCGGTCGCTGCTTTTTCATCCCTTGATCCCGCGGCTCCCTGCTGACTGTGC	481
Oy	481	CTGACGATCTGGAAATTAAGAGACCATGGCATCTGAGGTGCCCTTATGCCCTCTGGCATGCC	540
Db	482	CTGACGATCTGGAAATTAAGAGACCATGGCATCTGAGGTGCCCTTATGCCCTCTGGCATGCC	541
Oy	541	CATCAAGAAAATATGGSCATATGAAAGTGTGATTCCTCAGAGAGAGCAATATATAAAGAC	600
Db	542	CATCAAGAAAATATGGSCATATGAAAGTGTGATTCCTCAGAGAGAGCAATATATAAAGAC	601
Oy	601	AACCTCATACGCTCTTAAGAGGTGCCATCCAGTTAGGCAATTCCACACACTGTGGGAGCCT	660
Db	602	AACCTCATACGCTCTTAAGAGGTGCCATCCAGTTAGGCAATTCCACACACTGTGGGAGCCT	661
Oy	661	GAGTACCAAAACCGACGACGTGATGTCCCTCATGCGAATAATTTCTACGTGTWAGATATCTT	720
Db	662	GAGTACCAAAACCGACGACGTGATGTCCCTCATGCGAATAATTTCTACGTGTGTAGATATCTT	721
Oy	721	CTTTCCCACTGAAGAGAGCAACCTGACCCCTGCTCATCATCAATGACTTTGCTTTCAA	780
Db	722	CTTTCCCACTGAAGAGAGCAACCTGACCCCTGCTCATCATCAATGACTTTGCTTTCAA	781
Oy	781	GACCTATGACACCTGTTGGCCCTCCGCTACTTCCGGGAGGTATTTGGTATCCGGCCGATGA	840
Db	782	GACCTATGACACCTGTTGGCTTCGCGTACTTCCGGGAGGTATTTGGTATCCGGCCGATGA	841
Oy	841	TTACTGTATATTCCTCTCGCAGTGAAGCCGCTGATTAAGTCACTGTGAGCTCTGGAGCTAGTG	900
Db	842	TTACTGTATATTCCTCTCGCAGTGAAGCCGCTGATTAAGTCACTGTGAGCTCTGGAGCTAGTG	901
Oy	901	TTCCCTATCTATGTGTCCAGGGAGATGATGATTTATTTAAGACAGTCCAAACTAAGA	960
Db	902	TTCCCTATCTATGTGTCCAGGGAGATGATGATTTATTTAAGACAGTCCAAACTAAGA	961
Oy	961	GGCGGAATTTCTGCAGAAAGCTGCTTCCAGGATACACATGAACCTCAACAGAAACCTCG	1020
Db	962	GGCGGAATTTCTGCAGAAAGCTGCTTCCAGGATACACATGAACCTCAACAGAAACCTCG	1021
Oy	1021	GACTTTCCTGCTCAAAATTCATGTGACTGTGATGTGCGAGGCAAGGTGGCAAGAACATTGC	1080
Db	1022	GACTTTCCTGCTCAAAATTCATGTGACTGTGATGTGCGAGGCAAGGTGGCAAGAACATTGC	1081
Oy	1081	GATGTGTGATTAACAATATTTTACAGATCGGTAAAAAATGCATATCAATAATGTAGCT	1140
Db	1082	GATGTGTGATTAACAATATTTTACAGATCGGTAAAAAATGCATATCAATAATGTAGCT	1141
Oy	1141	CAAAAGCTCAACCTACAAACGCGCGGCTTCCAGAAAGGCGAGAGAAAGCCTCTTCCAC	1200
Db	1142	CAAAAGCTCAACCTACAAACGCGCGGCTTCCAGAAAGGCGAGAGAAAGCCTCTTCCAC	1201
Oy	1201	ATTTAAAGACCTGAGACTTCTTACAAACAATCCCTGATGCTTTTGTGATGCTGCAT	1260
Db	1202	ATTTAAAGACCTGAGACTTCTTACAAACAATCCCTGATGCTTTTGTGATGCTGCAT	1261
Oy	1261	GTACACGCTCTCTGTAGAGACCTGCGAGCGTGAAGTGTGTGCTGCGAGCCTTCAGAT	1320
Db	1262	GTACACGCTCTCTGTAGAGACCTGCGAGCGTGAAGTGTGTGCTGCGAGCCTTCAGAT	1321
Oy	1321	AATGATTAACAGCCTCTTGATGTGCATTCATATATATAGTATCATGTGACAAAGAGCCCTT	1380
Db	1322	AATGATTAACAGCCTCTTGATGTGCATTCATATATATAGTATCATGTGACAAAGAGCCCTT	1381
Oy	1381	AAGCAGTGAACACAGTACACTGTATACTCGAAGACCGCGCCCCCAAAAGGCTCTGTA	1440
Db	1382	AAGCAGTGAACACAGTACACTGTATACTCGAAGACCGCGCCCCCAAAAGGCTCTGTA	1441
Oy	1441	TTCCACACCCATGGATTCATCCAGGAGAGGCTCGACGGGTGCTACATGTGAGACTGA	1500

Db	1442	TTCCACAGCCATGGAAATCCATCCAGGGAGAGGCTCGACGGGGTGGTACCATGGAGACCTGA	1501
Qy	1501	TGACCATATGGGTGGCATCCCTGCGCGGAATAGTAAGGGGAAGGCTTCTGCTTTATAT	1560
Db	1502	TGACCAATATGGGTGGCATCCCTGCGCGGAATAGTAAGGGGAAGGCTTCTGCTTTATAT	1561
Qy	1561	TGGCATCATTTGACATCTCTACAGTCTTACAGGTTGTAAAGAAAGTTGGAGCACTCTTGGAA	1620
Db	1562	TGGCATCATTTGACATCTCTACAGTCTTACAGGTTGTAAAGAAAGTTGGAGCACTCTTGGAA	1621
Qy	1621	AGCCCTGGTATACATGACGAGACACTGTCTCTAGTGATGCAATGCGCCAGGCTTCTACGTGAACG	1680
Db	1622	AGCCCTGGTATACATGACGAGACACTGTCTCTAGTGATGCAATGCGCCAGGCTTCTACGTGAACG	1681
Qy	1681	GTTCACAGCGCTTCAATGTGCAACAGATTAAGATATCCCTCTGAAAGCCCTTCCTCTC	1740
Db	1682	GTTCACAGCGCTTCAATGTGCAACAGATTAAGATTAAGAAAGATTTCCCT	1723
Qy	1741	CAAAAAGTTTCGGTCTGGCTCATCTTCTCTCGGCGAGAGAGCTCCAGTGCMACTCCTG	1800
Db	1726	-----	1723
Qy	1801	CATTACTTACCAAGCATCGGTCTCTGGGGAAACAACAAGCAAGTACACAACAAAGCAGA	1866
Db	1726	-----	1723
Qy	1861	AGTGAGCCAGCGGCTTACCTTGGTGTGCTCGATGTTTACCTACAGACTCAACCTTTGGA	1920
Db	1726	-----GCGTTACACCTTGGTGTGCTCGATGTTTACCTACAGACTCAACCTTTGGA	1774
Qy	1921	GGAATCAGTGAAGGCTCGCTATTTCTGACCCCAAGTTTCTACCTCTAGTGGAGAAGAC	1988
Db	1775	GGAATCAGTGAAGGCTCGCTATTTCTGACCCCAAGTTTCTACCTCTAGTGGAGAAGAC	1834
Qy	1981	TTTGCAATGCTACTACAGTACAGTACAACTTGGAAAAGCTGAAGTGCAGAGTCAAGATT	2040
Db	1835	TTTGCAATGCTACTACAGTACAGTACAACTTGGAAAAGCTGAAGTGCAGAGTCAAGATT	1894
Qy	2041	CACCCATTAAAGCGCAAAAGCCTCAGAGACCTGGAAACAAGATTGCTGCTATCTGTGATCC	2100
Db	1895	CACCCATTAAAGCGCAAAAGCCTCAGAGACCTGGAAACAAGATTGCTGCTATCTGTGATCC	1954
Qy	2101	CAAGATGTCAAGCCCTTGCGCCAGCAATGCTGAATTTTCTCTACTGATGATCAAA	2156
Db	1955	CAAGATGTCAAGCCCTTGCGCCAGCAATGCTGAATTTTCTCTACTGATGATCAAA	2010
RESULT 4			
AA573985			
ID	AA573985	standard; cDNA; 1992 BP.	
XX	AA573985;		
AC			
XX	13-FEB-2002	(first entry)	
DE			
XX		DNA encoding novel human diagnostic protein #9789.	
XX			
KW		Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW		food supplement; medical imaging; diagnostic; genetic disorder; ss.	
OS		Homo sapiens.	
XX			
PN		WO200175067-A2.	
XX			
PD		11-OCT-2001.	
XX			
PF		30-MAR-2001; 2001WO-US08631.	
XX			
PR		31-MAR-2000; 2000US-0540217.	
PR		23-AUG-2000; 2000US-0649167.	
XX			
PA		(HYSE-) HYSEQ INC.	

	RESULT	5
ID	AAT84358	
	AAT84358 standard; DNA; 2764 BP.	
XX		
AC	AAT84358;	
XX		
DN	12-NOV-1997 (first entry)	
XX		
DE	STM7.I gene associated with Friedreich's ataxia.	
XX		
KW	STM7.I gene; Friedreich's ataxia; FRDA; neurodegeneration; phosphatidylinositol-4-phosphate 5-kinase; PtdInsPK; transgenic animal; animal model; diagnosis; therapy; ss.	
KX		
OS	Homo sapiens.	
XX		
FH	Key Location/Qualifiers	
FT	CDS 420..2042	
FT	/tag= a	
FT	PolyA_signal 2749..2753	
FT	/tag= b	
XX		
PN	M09705Z34-A2.	
PD	13-FEB-1997.	
XX		
PF	24-JUL-1996; 96WO-GB01786.	
XX		
PR	28-JUL-1995; 95GB-0015508. 26-JUL-1995; 95GB-0015309.	
PA	(UNLO) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.	
PI	Chamberlain S, Doudney CME, Garcia-Valdecasas JJC; Hillermaier R, Pook MA;	
DR	WPJ: 1997-145674/L3. P-PSDB; AAM00978.	
PT	New gene, STM7, and new protein causing Friedreich's Ataxia - encodes phosphatidylinositol phosphate kinase isoform; also splice variants and transgenic animals, for diagnosing pre-disposition to Friedreich's ataxia	
PS	Claim 1; Fig 7; 82pp: English.	
CC	This isolated DNA molecule comprises the full-length human STM7.I gene associated with Friedreich's ataxia (FRDA). It comprises exons 1-16 of the STM7 gene (see AAT84364-86) and encodes a phosphatidylinositol-4-phosphate 5-kinase (AAM00978) isoform (P5K).	
CC	The STM7.I sequence was obtained by construction of chromosome 9 region-specific cosmid contigs, exon trapping, infant cerebellum cDNA library screening and RACE. STM7 nucleic acids can be used to produce recombinant p5k in recombinant host cells using a claimed method. Probes derived from exons 1-16, partic. 7-11, of STM7 are used in a claimed diagnostic method for determining an inherited predisposition to FRDA. STM7 nucleic acids can also be used to generate transgenic animal models of FRDA and in somatic cell therapy.	
SQ	Sequence 2764 BP; 850 A; 567 C; 573 G; 774 T; 0 other:	
Query Match	15.7%; Score 581.6; DB 18; Length 2764;	
Best Local Similarity	69.6%; Pred.No. 3.2e-153;	
Matches 820; Conservative	0; Mismatches 349; Indels 9; Gaps 2	
OY	578 GGAGGACAACATTAAAAAACAACCTCATCAGCGTTGAAGAAGTCATCAAGTAAAGGC 637 	
Db	468 GAAGAAAAAACCTTAATAAAAAAGACTGATCATCTGGCATTAAGAAGGTGCTATTCAGCTGGGA 527	
OY	638 ATTACCACACAGTGTTGGGGAGCTGAGTACCAACAACAGACGAGTAGTCTCATGCCAAGAT 697 	
Db	528 ATAGATATCACAGTGCGGTATCTCACTTCACAGCCAGAAACGAGATGTTTATATGCAAAGAC 587	

QY	698	TTTCACGTGGTGAAGTATCTCTTTCCACGAGGAGGACAACTGACCCCTCAT	757
Db	588	TTTTATGTGGTGAAGTGTGTCTTCACCCAGGAGGAGGACATCTGACCCACACAT	647
QY	758	CACACAAATGACTTGGTTTTCAAGACCTATGACACTGTTGCCCTCCGCTACTTCCGGGAG	817
QY	818	CTATTGTGTAACGGGCCGATGATTCCTGTATTCCTCGACAGGAGGCGGTGATTAA	877
Db	708	CTTTTGGTATCAAGCCTGTAGATATCTGTATATTCATCTGGCAGTGAACCTTAAATGAA	767
QY	878	CTCTGTAGCTCTGGAGCTAGCTGGTCCCTATTCCTATGTGTCCAGCGACGATAGTTCAAT	937
Db	768	CTGTCTAACCCCTGGAGCCAGTGGATCCTGTTTGTGTGACAGATGATGATTAATTTATC	827
QY	938	ATTTAAGCAGTCCACACTTAAGAGCGGGAATTTCTGCAGAGCTGCTTCCAGATACTAC	997
Db	828	ATCAAAAACAGTTCAGCCCAAGAAAGGTGAATTTCTTCAGAACGTACTGCGACAGCTATTAC	887
QY	998	ATGAACCCCAACCGAACCCCTGGAGCTTTCGCTCAATTCGATAGGACTGTACGTGTG	1057
Db	888	ATGAATTTTAAACCAAGATCCAGAGACTCTTTTCCAAAATTTTACGGACTGTATGTATG	947
QY	1058	CAGGAGGTGGCAAGAACATTGGATTTGTGTGATGAACAAATCTTTTACCAGATCGGTA	1117
Db	948	CAATCAGAGAGCATTAATATTCAGAAATGTGTGATGACAAAGTTTGGCAGCTGCATG	1007
QY	1118	AAATATGATNTCAAAATNTGACCTCAAAAGGTCAACCTACAAAACGCGGGCTTCCAGAA	1177
Db	1008	AGAAATGCACTTTACATNTGACTTGAAGGCTCAAGCTAATTAAGCGAAGACATCCGTAA	1067
QY	1178	GAGCGAGAGAAGCCCTCTCCACATTTTAAAGACCTAGACTTCTTACAAAGACATCCCTAT	1237
Db	1068	GAGAGAGAAATCCAAACCCACATTTTAAAGACTGTAGATTTCTGTGAAGACATGACGA	1127
QY	1238	GGCTCTTTTGGATGCTGACATGTACAAAGCTCTCTGTAAAGCCCTGCAGCTGACTGT	1297
Db	1128	GGGTGTATTTTGTATGTAAGAAATCATCAAGCGCCTTATGAAAACACTTTCAGAAACATGC	1187
QY	1298	TTTGCTGTGAGAGCTTCAAGATTAATGATTAAGAGCTCTGTAGTGCATTCATTAATTA	1357
Db	1188	CGGGTGTCAAAAAGCTTCAAGATCATGATGATTAATGAGCTTCTGTGGGAATTAATTTCTG	1247
QY	1358	GATCATGCACACGAGAGCCCTTAAGCAGTGAACACAGTACTCAAGTGTACTGAAAGA	1417
Db	1248	GACCAATTCCTCAAGAGAAAGAGAGAGAGACCCTCAAAAATGTGCTGATCTAAAGCG	1307
QY	1418	CCGGCCCCCAAAAGGCTGTATTCACAGCCATGGAATTCATCCAG-----GGAAG	1477
Db	1308	ACTGGATGCAAGAGGTGTCTACTCAACAGCCATGGAATCTATCCAGGGTCCAGGAAA	1367
QY	1472	GCTCGACGGGGTGTACCATGAGAGACTGATGACATATATGGGTGGCACTCCCGGAT	1537
Db	1368	TCCTGAGATGGGATTAATCAACAGAGAACCCAGACAAATGGGAGGACTTCCAGCTAAAGC	1427
QY	1532	AGTAAAGGGGAAAGGCTTCGCTTATATTTGGCATCTATGACATTTCTACAGTCTTAAG	1597
Db	1428	CATGGGGGAGAAAAACATATTTATTTATGGGATTAATGAACTCTGCAATCATATAGG	1487
QY	1592	TTTGTTAAGAGTGGAGCACTCTTGGAAGAGCCCTGGTCAATGAGGAGACACTGTGCA	1657
Db	1488	TTAATGAAGAGTTAAGAACTTCTCTGAAAGCTCTGTATTATGATGGGACACTGTCTT	1547
QY	1652	GTGCAATCGCCCAAGCTTCTACGCTGAACGGTTTCAGCGCTTCAAGTGCACACAGTATT	1717
Db	1548	GTTCATAGACCAAGCTTTATATGACAGACAGATTTCTTAAGTTCATGAATCCAGAGATTTC	1607
QY	1712	AAGAAGATC---CCTGAAGCCTTCCCTCCAAA	1746
Db	1608	AAGAAATTCACGCTTGAAGGCTTCAACGCTCAAG	1645

SQ Sequence 598 BP; 149 A; 156 C; 139 G; 154 T; 0 other;

Query Match	12.7%;	Score 469.8;	DB 22;	Length 598;
-------------	--------	--------------	--------	-------------

Matches 527; Conservative 0; Mismatches 52; Indels 6; Gaps 2;

QY	1331	AGCCTCTGATGTCATCCATTAATATAGATCATGCACAACGAGAGAGCCCTTAAGCAGTAA	1390
QY	1331	AGCCTCTGATGTCATCCATTAATATAGATCATGCACAACGAGAGAGCCCTTAAGCAGTAA <td>1390</td>	1390
Db	2	AGCCTCTGAGCTGTCAATCCACATATATAGATCATGCACAACGAGAGAGCCCTTAAGCAGTAA	61
QY	1391	AC--AAGTACTGACTGTGATTAAGTACTGAGACGCGCCGCCCAAAAGAGCTCTGATTCCACA	1447
Db	62	ACTCTTAAATGTCTCAACGACACTCAAAAGCTGGCTCCCAAAAGAGCTCTGATTCCACA	121
QY	1448	GCCATGGAATTCATCCAGGAGAGAGGCTCGACGGGGGTGTACATGAGAGCTGATGACAT	1507
Db	122	GCCATGGAATTCATCCAGGAGAGAGGCTCGGCTGGGGAACCATGTGAGGCCGATATACAT	181
QY	1508	ATGGGTGGCATCCCTGCGCGGAATAGTAAAGGGGAAAGGCTTCTGCTTATATTTGGCATC	1567
Db	182	ATGGGTGGCATCCCTGCTCGAATAGTAAAGGGGAAAGGCTTCTGCTTATATTTGGCATC	241
QY	1568	ATTTACATTTTACAGTCTTTACAGTCTTTGTTAAGAAAGTTGGACACGCTTGGAAAGCCCTG	1627
Db	242	ATTTACATTTTACAGTCTTTACAGTCTTTGTTAAGAAAGTTGGACACGCTTGGAAAGCCCTG	301
QY	1628	GTCATATACGAGACACTGTCTCATGTCATCGCCACAGGCTTCTACGTGAACGTTCCAG	1687
Db	302	GTCATATATATGGGAGCGGTGCTCATGTCATCGCCACAGGCTTCTACGTGAACGTTCCAG	361
QY	1688	CGCTTCATGTGCAACACAGATTTTAAAGAAATTCCTTCAACGCTTCTCCCAAAAG	1747
Db	362	CACCTTCAATGTGCAACGAGATTTTAAAGAAATTCCTTCAACGCTTCTCCCAAAAG	421
QY	1748	TTTTCGCTGCGCTCATCTTTCTCTCGGAGACGAGGCTCCAGTGGCACTCTGCAATTACT	1807
Db	422	TTTTCGCTGCGCTCATCTTTCTCTCGCATACGAGGCTCCAGTGGCACTCTGCAATTACT	481
QY	1808	TACCAAGCATCGGCTCTGTGGGAAACAAAGGCACAAAGTGACAACAAAGGCAAGTGGAG	1867
Db	482	TACCAAGCATCTGTCTGTGAGAAACAAAGTCAACAAAGT--ATTAAGGTGCAAAGTGGAG	538
QY	1868	CCAAGCGTTCACCTTGGTGTCTGTATGTTTAACTCAGACTCA	1912
Db	539	CCAAGTGTTCACCTTGGTGTCTGTATGTTTAACTCAGACTCA	583
RESULT 8			
AAK36262			
ID	AAK36262 standard; DNA; 598 BP.		
XX			
AC	AAK36262;		
XX			
DT	06-NOV-2001 (first entry)		
XX			
DE	Human bone marrow expressed single exon probe SEQ ID NO: 10819.		
XX			
KW	Human; bone marrow expressed exon; gene expression analysis; probe;		
KW	microarray; cancer; Leukemia; Lymphoma; myeloma; ss.		
OS	Homo sapiens.		
XX			
PN	W0200157276-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	30-JAN-2001; 2001WO-US00668.		
XX			
PR	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		

Query Match	Best Local Similarity	Score	DB	Length
Matches 527; Conservative 0; MissMatches 52; Indels 6; Gaps 2;	12.7%;	459.8;	DB 22;	598;
Query 1331	AGCCCTTGATGTCATCATCATATATATAGTCAATGCAACAAGAGAGCCCTTAAAGCATGAA	1390		
Db 2	AGCCCTTGATGTCATCATCATATATAGTCAATGCAACAAGAGAGCCCTTAAAGCATGAA	61		
Query 1391	AC--ACAGTACTGATGTTGATCTGGAAGACCGCCGCCCAAAAGGCTCTGATTCCACA	1447		
Db 62	ACTCTTCAAGTGTCAATGCACACTCAAGAGCTGCTCCCAAAAGGCTCTGATTCCACA	121		
Query 1448	GCCATGGAAATCATCAAGAGAGAGGCTGACGCGGCTGATCATGGAACATGATGACAT	1507		
Db 122	GCCATGGAAATCATCAAGAGAGAGGCTGACGCGGCTGATCATGGAACATGATGACAT	181		
Query 1508	ATGGGTGGCATCCCTGCGCCGGAATAGTAAAGGGGAAAGGCTCTGCTTAATATTTGGCATC	1567		
Db 182	ATGGGTGGCATCCCTGCGCCGGAATAGTAAAGGGGAAAGGCTCTGCTTAATATTTGGCATC	241		
Query 1568	ATTGACATTCTACATCTTTACAGGTTTGTTAAGAACTTGGACACTCTTGGAAAGCCCTG	1627		
Db 242	ATTGACATTCTACATCTTTACAGGTTTGTTAAGAACTTGGACACTCTTGGAAAGCCCTG	301		
Query 1628	GTACATGACGAGACACTGCTCACTGATGCCAGGCTTCTAGCGTGAAGGTTCCAG	1687		
Db 302	GTACATGATGGGAGACGCTGCTCATGATGCCAGGCTTCTAGCGTGAAGGTTCCAG	361		
Query 1688	CGCTTCATGTCACACAGATATTTAAGAAAGATTCCTTGAAGCCTTCTCTCCAAAAAG	1747		
Db 362	CACCTTCATGTCACACAGATATTTAAGAAAGATTCCTTGAAGCCTTCTCTCCAAAAAG	421		
Query 1748	TTTGGTGTGGGTATCTTCTCTGCGGAGAGAGGCTCCAGTGGGAACCTCGCATTTCT	1807		
Db 422	TTTGGTGTGGGTATCTTCTCTGCGGAGAGAGGCTCCAGTGGGAACCTCGCATTTCT	481		
Query 1808	TACACGACCATGCTCTCTGGGGAACACAAAGGACAAAGTACAAACAAAGGACAGATGGAG	1867		
Db 482	TACACGACCATGCTCTCTGGGGAACACAAAGGACAAAGTACAAACAAAGGACAGATGGAG	538		
Query 1868	CCAGGCGTTACACCTTGCTGCTCTGATGTTTAACTCAGACTCCA	1912		
Db 538	CCAGGCGTTACACCTTGCTGCTCTGATGTTTAACTCAGACTCCA	583		
Result 9	AA141984			
ID	AA141984	standard; DNA: 598 BP.		
CC	AA141984;			

[illegible]

Db 971 ACGATTTCCTCCGCCAGAGGTTCTCCATACACCTGCCCATCTACACAGCTTC 1030
 Qy 773 GCTTCAAGACCTATGACACTGTTGCTTCCGCTACTTCCGGAGACTATTTGATCCGG 832
 Db 1091 CGCTCAAGGCTACAGACCCATGCTTCCGCTACTTCCGAGACTCTGCGAGTCCGA 1090
 Qy 833 CCGGATGATTAATGTTCCCTCTGACAGTGAAGCCGCTGATGAACCTGATCTCTGGA 892
 Db 1091 CCGGATGATTAATGTTCCCTCTGACAGTGAAGCCGCTGATGAACCTGATCTCTGGA 1150
 Qy 893 GCTAGTGGTCCCTATGATGTCACAGAGTGGTCCGATGATTAATGAAGACTCCAA 952
 Db 1151 GCTAGTGGTCCCTATGATGTCACAGAGTGGTCCGATGATTAATGAAGACTCCAA 1210
 Qy 953 CATAAAGAGGCGAATTTCTGACAGAGTGGTCCGATGATTAATGAAGACTCCAA 1012
 Db 1211 AAGAAGAGAGTGGTCTCTGCAAAACTTTGGCCGTTATATAGACTATGCGAG 1270
 Qy 1013 AACCTCGGACTTGTCTGCTTAATTTCTATGAGACTGTACTGTGCGAGGCGAGTGGCAAG 1072
 Db 1271 AATCTCGAGAGCTGCTGCTTAAGTTCTTCCGACTATATGCTTCCACTACAACTCGAAG 1330
 Qy 1073 AACATGAGATTTGGTGTATGACAACTTTTACCAAGATGCGTAAATGATATCAAA 1132
 Db 1331 AACGTCGGTGTGGTCCATGACAACTGTTGCTTACAGACTTAAGATGCACTGCAAG 1390
 Qy 1133 TATGACTCTCAAGGCTCAACTACAAAGCGGCGCTTCCGAAAGAGGAGAGAGGCT 1192
 Db 1391 TACGACTCTTAAGGATCTTCTCTCGAGAGGAGGCTTCCAAAGGAGAGGAGGAGGCG 1450
 Qy 1193 CTTCACCATTTTAAAGACCTAGACTTTCTTACAAAGCATCCCTGATGATCTTTTGGAT 1252
 Db 1451 AGTCCACGTTCAAGAGCCTGACTTGTCTGAGCAGCAATCCGAATGGATATTTCTGGAG 1510
 Qy 1233 GCTGACATGTACAGGCTCTCTGTAAAGCCTGCGAGCTGACTGTTGGTGTGCGAGAGC 1312
 Db 1511 ACGGACGATCAATGACACTGATGAGCAGCAATTAAGCGGAGCTGATGCTGTGAGAGT 1570
 Qy 1313 TTCAAGATATGATTAACGCTCTGATGTCATTCATTAATATATGATCATGC 1365
 Db 1571 TTTGAGATTAATGACTACTGCTGTTGGTGGAAATCCAAACCTGAGACTGGC 1623
 RESULT 14
 ABL15962/C
 ID ABL15962 standard; cDNA; 6699 BP.
 XX
 AC ABL15962;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 42368.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN W0200171042-A2.
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI; 2001-656860/75.
 DR P-PSDB; ABB71859.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX
 PS Claim 1: SEQ ID NO 42368; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB13511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB172072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 SO Sequence 6699 BP; 1663 A; 1551 C; 1541 G; 1944 T; 0 other;

Query Match 9.3%; Score 346.6; DB 23; Length 6699;
 Best Local Similarity 63.5%; Pred. No. 1.5e-86;
 Matches 529; Conservative 0; Mismatches 304; Indels 0; Gaps 0;

Qy 533 GGCATGCCCATCAAGAAATAGCCATAGAAAGTTGATTTCTCAGAGAGACAATAT 592
 Db 3519 GGGAAAGCAGAAAGAGAGTGGCCAGCCAGAGTGGCGAGAGGCGGTGAGTGCATAC 3460
 Qy 593 AAAAAAGCAACCTCATCAGCCTTGAAGAGTGCATTCAGATTAAGCATACCAACATGTG 652
 Db 3459 AAGAAGATCAATCAAGCAAAATATGAGCTCATTTACAGCTGGGATCCAGATACGGTG 3400
 Qy 653 GGGAGCTGATGACCAAGAGAGTGTATGTCATGACCAAGATTTCTACGTGGTGA 712
 Db 3399 GGCAGCCTGGCCAGTAAGCCCAAGGCGGATGCTTAATGAAGACTTCTGGAAATGGA 3340
 Qy 713 AGTATCTTCTTCCAGTGAAGGAGCAACTGACCCGCTCATCACTCAATGATGAT 772
 Db 3339 AGGATTTGCTTCCGCGAGATGTTCTTCATACACCTGCCCATACACAGCACTTC 3280
 Qy 773 GCTTCAAGACCTATGACACTTGTGCTTCCGCTACTTCCGAGAGCTATTTGGATCCGG 832
 Db 3279 CCGTTCAAGGTCTAGACACCTATCCGCTTCCGCTACTTCCGAGATCTGTTCCGAATCGCA 3220
 Qy 833 CCGGATGATTAATGATGATCCCTGACAGTGAAGCCGCTGATGAAGCTGTAGCTCTGA 892
 Db 3219 CCGGATGATTAATGATGATCCCTGACAGTGAAGCCGCTGATGAAGCTGTAGCTCTGA 3160
 Qy 893 GCTAGTGGTCCCTATGATGTCACAGAGTGAAGTGAATTAATGAAGACTCCAA 952
 Db 3159 GCTCTGTGTCATATTTTACTGACAGCAGAGAGTTCATATTAAGAGCGGTGCAAG 3100
 Qy 953 CATAAAGAGCGGAAATTTCTGACAGAGTGGTCCGATGATTAATGAAGACTCCAA 1012
 Db 3099 AAGAAGAGTGGAGTCTCTGCAAAACTTTGGCCGTTATATAGACTATGCGAG 3040
 Qy 1013 AACCTCGGACTTGTCTGCTTAATTTCTATGAGACTGTACTGTGCGAGGCGAGTGGCAAG 1072
 Db 3039 AATCTCGAAGCCTGCTGCTAAGTTCTTCCGACTATATGCTTCCACTCAACATCGAAG 2980
 Qy 1073 AACATGAGATTTGGTGTATGAACAATCTTTTACCAAGATGCGTAAATGATATCAAA 1132
 Db 2979 AACGTCGGTGTGGTCCATGACAACTGTTGCTTACAGATTAATGAAGTGCATGCAAG 2920
 Qy 1133 TATGACTCTCAAGGCTCAACTACAAAGCGGCGCTTCCGAAAGAGGAGAGAGGCT 1192
 Db 2919 TACGACTCTTAAGGATCTTCTCGAGAGGAGGCGCTCAAGAGGAGGAGCAAAAGGCG 2860
 Qy 1193 CTTCACCATTTTAAAGACTTACTTCTTCAAGAGCATCCCTGATGATCTTTTGGAT 1252
 Db 2859 AGTCCACGTTCAAGAGCCTGACTTCTGTGAGCAGCAATCGAATGCGATATTTCTGGAG 2800

Qy	2007	CCGGGAAAAGCTTTGAAGTTGGCAGAGTCAGAGTTACCCATTAAACGCAAAAGCCCTCAGAA	2066
Db	313	CTTTGGAAAAGCTTTGAGTTGGCAGAGTTCAGAGTTACCCATTAAAGCGCAAAAGCCCTCAGAA	254
Qy	2067	GACCTGGACAAAGATTTCTGCCATCTCTGTATGCCAAGATGTACAGCCCTTGCCCCAGCA	2128
Db	253	GACCTGGACAAAGATTTCTGCCATCTCTGTATGCCAAGATGTACAGCCCTTGCCCCAGCA	194
Qy	2127	TGCTGAATTTTCTTCTACTTGGTCATCAAAAAAGAGAGTAAATAGAAGAGAGGAGCTG	2186
Db	193	TGCTGAATTTTCTTCTACTTGGTCATCAAAAAAGAGAGTAAATAGAAGAGGAGGAGCTG	134
Qy	2187	CTCTCCATCTTCTTCTCTGAAGAAGAACTTCTCTCTCTCTCTCTCTCTCTATGAATAGGGCC	2246
Db	133	CTCTCCATCTTCTTCTCTGAAGAAGAACTTCTCTCTCTCTCTCTCTCTCTATGAATAGGGCC	74
Qy	2247	TTATGTGCTCAGAGAGTTGAGGACCGAGCAGTCCCTCCACTCCAGAGTTGGGTGTACG	2306
Db	73	TTATGTGCTCAGAGAGTTGAGGACCGAGCAGTCCCTCCACTCCAGAGTTGGGTGTACG	14
Qy	2307	GATTTTCAACATCG 2319	
Db	13	GATTTTCAACATCG 1	

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

1      PRIOR APPLICATION NUMBER: US 09/774,203
2      PRIOR FILING DATE: 2001-01-29
3      NUMBER OF SEQ ID NOS: 49117
4      SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
5      SEQ ID NO 13058
6      LENGTH: 598
7      TYPE: DNA
8      ORGANISM: Homo sapiens
9      FEATURE:
10     OTHER INFORMATION: MAP TO AL096800.20
11     OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.72
12     OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.53
13     OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.53
14     OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.65
15
16 US-09-864-761-13058

```

Query Match	12.7%;	Score 469.8;	DB 10;	Length 598;
Best Local Similarity	90.1%;	Pred. No. 2e-119;		
Matches 527;	Conservative 0;	Mismatches 52;	Indels 6;	Gaps 2.

OY	1331	AGCCCTTGATGTCATATCCCAATTAATATGATCATGCACAACGAGAGCCCTTAAGCAGTAA	1390
Db	2	AGCCTCTGGCTGTCAATCCCAATATATGATCATGCACAACGAGAGCCCTTAAGCAGCAG	61
OY	1391	AC--ACAGTACTCAGTTGATTACTCGAAGACCGGCCCCCAAAAGGCTCTGTATTCCACA	1447
Db	62	ACTCTCAAGTGTCAATTCGAATCGAAGCTCAAAAGACTGGCTCCCAAAAAGCTCTGTATTCCACA	121
OY	1448	GCCATGGAATTCATTCATCCAGGGAGAGGCTTCACAGGGGGGTGTATCCATGGAGACTATATACAT	1507
Db	122	GCCATGGAATTCATTCACAGGGAGAGGCTTCAGGCTGGGGACACCAATGGAGGCGCATATACAT	181
OY	1508	ATGGGTGGCATCCCTGGCCGGGAATAGTAAGAAAGGGGAAAAGGCTTCTGCTTAATATTGGCATC	1567
Db	182	ATGGGTGGCATCCCTGCTCAGAAATAGTAAGAAAGGGGAAAAGGCTTCTGCTTAATATTGGCATC	241
OY	1568	ATTGCATTTTACAGTCTTACAGGTTTGTTAAGAAGTTGGAGCATCTTTGGAAAGCCGTG	1627
Db	242	ATTGCATTTTACAGTCTTACAGTCTTACAGTCTTAAAGAAGTTGGAGCATCTTTGGAAAGCCGTG	301
OY	1628	GTACATGACGGAGACACTGTCTCAGTGTCCATTCGCCAGGCTTCTACGCTTAAGGTTTCCAG	1687
Db	302	GTACATGATGGGGAGCGCTGTCTCAGTGTCCATTCGCCAGGCTTCTACGCTTAAGGTTTCCAG	361
OY	1688	CGCTTCATGTGCAACACAGATATTTAAAGAATTCCTTGAAGGCTTCTCTCCATCAAAAAG	1747
Db	362	CACCTTCATGTGCAACGCGATATTTAAAGAATTCCTTGAAGGCTTCTCTCCATCAAAAAG	421
OY	1748	TTTCGGTCTGAGCTATCTTTCTCTCGGCGAGCAGGCTCCAGTGGCAACTCTGTCAATTACT	1807
Db	422	TTTCGGTCTGAGCTATCTTTCTCTCGCTCATACGGGCTCCAGTGGCAACTCTGTCAATTACT	481
OY	1808	TATCAGCCATGCGTCTCTGGGGAAACACAAGGCACAACTGACAAACAAGCAGAACTGGAG	1867
Db	482	TATCAGCCATGCGTCTCTGAGGAACACAAGTCACAAAGT---ATTAAGGTGCAAGTGGAG	538
OY	1868	CCAGGCGTTCACCTTGGTCGTCGTCGATAGTTTACTCAGACTCCA	1912
Db	538	CCAGGCGTTCACCTTGGTCGTCGTCGATAGTTTACTCAGACTCCA	583

```

RESULT 3
US-09-954-456-827
; Sequence 827, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954, 456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617

```


QY 3643 GTTCAGAAATGCAATACATATAAAGTATATGTTTAATGTAATAACTTTAA 3702
DB 72 GTTCAGAAATGCAATACATATAAAGTATATGTTTAATGTAATAACTTTAA 13
QY 3703 TGAGTTATTTTA 3713
DB 12 TGAGTTATTTTA 2

RESULT 5
US-09-864-761-29620
; Sequence 29620, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29620
; LENGTH: 346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL096800.20
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.72
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.53
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.53
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.65
; OTHER INFORMATION: SWISSPROT HIT: P38994, EVALUE 4.00e-07

OTHER INFORMATION: NT HIT: g14505814, EVALUE 1.00e-125
OTHER INFORMATION: EST_HUMAN HIT: BE794576.1, EVALUE 1.00e-125
US-09-864-761-29620

Query Match 7.7%; Score 285; DB 10; Length 346;
Best Local Similarity 92.3%; Fred. No. 1.3e-68;
Matches 300; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1400 TCAGTTGATCTGCAAGACCGCCGCCCAAGGCTGTATTCACAGCCATGATCC 1459
DB 22 TCATATGACACTCAAGAGCTGCTCCCAAGGCTGTATTCACAGCCATGATCC 81
QY 1460 ATCCAGGAGAGAGCTCGACGGGGTGTACCATGACATGATGACATGATGATG 1519
DB 82 ATCCAGGAGAGAGCTCGCTGGGCGCACCATGAGGCGCATGATGATGATGATG 141
QY 1520 CCTGCCCGAATAGTAAGGAGGAAGGCTTCGCTTATATGATGATGATGATGATG 1579
DB 142 CCTGCTCAGAAATGTAAGGAGGAAGGCTTCGCTTATATGATGATGATGATGATG 201
QY 1580 CAGTCTTACAGGTTTGTATTAAGAGTTGAGACACTCTTGGAAAGCCCTGTACATGACGGA 1639
DB 202 CAGTCTTACAGGTTTGTATTAAGAGTTGAGACACTCTTGGAAAGCCCTGTACATGATGATG 261
QY 1640 GACACTGTCTCAGTGCATGATGCCAGGCTTCTACGCTGAACGCTTCAGGCTTATGTC 1699
DB 262 GACGCTGTCTCAGTGCATGATGCCAGGCTTCTACGCTGAACGCTTCAGGCTTATGTC 321
QY 1700 AACACAGTATTAAGAAGATTCCT 1724
DB 322 AACGCAGTATTTAAGAAATCCCT 346

RESULT 6
US-09-783-590-9100
; Sequence 9100, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9100
; LENGTH: 350
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (24)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (40)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (44)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (134)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (172)
; OTHER INFORMATION: n equals a,t,g, or c


```
NAME/KEY: misc feature
LOCATION: (173)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (184)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (189)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (193)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (203)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (209)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (217)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (245)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (250)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (256)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (282)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (298)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (317)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (321)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (323)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (350)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-9100
```

```
Query Match 4.1%; Score 152; DB 10; Length 350;
Best Local Similarity 69.9%; Pred. No. 5.9e-32;
Matches 232; Conservative 0; Mismatches 95; Indels 5; Gaps 3;
```

```
QY 2448 GGACAGCTAGCTGGACATTAATCCCTAGTTCCTCTCTGACT-CCGGAAGA 2506
DB 9 GAACAGACTAGCTGCNACATATATCCCACTNAGTCTCTGAACTCCGGAAGA 68
QY 2507 ATACTCCGTATCTCTGTAAGGTTTGGGGGTAAGGGTTTAACCACTCCGAC 2566
DB 69 AAACCCCGTATCTCTGTAAGGTTTGGGGGTAAGGGTTTAACCACTCCGCA 128
QY 2567 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2626
DB 129 CTT-TGNTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 187
QY 2627 ATTTTCAATCAGAACTCCAGAGTTGACAGATGCCATTCGTAGAGTTCCCTCAGA 2686
DB 188 ANTITNAGTCAAAANTCCAGNGGTTTGGGGGTTGCTTCCCTGGGATTCCTC 247
QY 2687 AGAG---CCATGCTGTATATGAGAGAGAGATGATTCCTCTCCGAGAACAGCTCCT 2743
DB 248 AANGGCCCGGTTTAAAGGGAAGGTGGGTTGCTCTCCCAANAGGATTCCTC 307
```

```
QY 2744 CTTTAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2775
DB 308 TTTTAAATNCCNANTTTTGGGGAATTTT 339

RESULT 7
US-09-938-842A-2662
Sequence 2662, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kieps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938, 842A
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227, 866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264, 647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300, 111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 2662
LENGTH: 2580
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-2662
```

```
Query Match 2.0%; Score 75; DB 9; Length 2580;
Best Local Similarity 49.2%; Pred. No. 3.7e-10;
Matches 325; Conservative 0; Mismatches 315; Indels 21; Gaps 4;
```

```
QY 716 ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 775
DB 1426 ATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1485
QY 776 TTCAAGACCTATGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 835
DB 1486 TCGAAGACTATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1545
QY 836 GATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 892
DB 1546 GAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1605
QY 893 GCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 952
DB 1606 AAAAGTGGAGATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1665
QY 953 CATTAAGAGCGGAAATTTCTGAGAAAGCTGTTCCAGATGATGATGATGATG 1012
DB 1666 AAGTCTGAGTTGACAGTTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1725
QY 1013 AACCTCGGACTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1069
DB 1726 CATGAAACACACTTATTAACCAAAATTTTGGAGTTCACAGATTAACCTCA 1785
QY 1070 AAGAATCTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 1129
DB 1786 AAAAAGTACGCTTGTATCTATGAGGAAATATGTTTGCAGAAATGAAATG 1845
QY 1130 AAATATGACCTAAAGGCTCAACCTTCAAAAGCGGCGCTCCAGAAAGAGG 1189
DB 1846 CGTATGATCTAAAGGCTCAACCTTCAAAAGCGGCGCTCCAGAAAGAGG 1902
QY 1190 CCTCTCCCACTTAAAGACCTGAGACTCTTACAAGATCCCTGATGAGTCT 1249
DB 1903 GAGAAAGACCACTTGAAGATCT-----TGATCTAGCTTATGAAATTCAT 1950
```

QY 1250 GATGCTGACATGACACGCTCTCTGTAGACCCCTGCAGCGTCTGTTGGTCTGAC 1309
DB 1951 GACACAGCTGTAGGGAGCCCTTTTCACCAATTTTACTTACCTCTGTTTGGAA 2010
QY 1310 AGCTTCAGATTAATGATTAACAGCTCTGTATGATCATCATATATATGATGACAA 1369
DB 2011 TCCTGTAACATTCATGACTACAGCTCTTTACTGGGATTAATTTTACAGCTCTCGCCAA 2070
QY 1370 C 1370
DB 2071 C 2071

RESULT 8

US-10-012-055-3
; Sequence 3, Application US/10012055
; Patent No. US20020164750A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: 56634, A NOVEL HUMAN PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE
; FILE REFERENCE: 10448-114001
; CURRENT APPLICATION NUMBER: US/10/012,055
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 60/248,325
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-055-3

Query Match 2.0%; Score 73.8; DB 9; Length 1266;
Best Local Similarity 48.8%; Pred. No. 5e-10;
Matches 297; Conservative 0; Mismatches 297; Indels 15; Gaps 3;

QY 756 ATCACTACATGACTTTCGTTCAAGACCTATGACCTGTGCTTCCGCTACTTCGGG 815
DB 278 ATCTGCCAGTCAATTCATGATCAAGAGATATGTCGCCAGTCTTCAGGAACTCCGTG 337
QY 816 AGCTATTGGTATCGGCCGATGATTAATCTGATTCCTCTGACAGAGCCCTGATTTG 875
DB 338 ATGATTTGGCATGATGACCAAGATTAATCTGATGCTTACCCGAAACCCGCCAGCG 397
QY 876 AACTCTGATGCTGTGAGCTAGTGTCTCTATTTATGATGTCGAGGAGCATGATTC 935
DB 398 AAGG-----TGAAGGAGTGAATGATGCTCTCTTATCTCTACGATCGGACTCTG 448
QY 936 TTATTAAGACATGCAACATTAAGAGCGGAATTTCTGCAAGAGCTGCTTCCAGATCT 995
DB 449 TCATCAAGAGATGATCAGAGGACATTCGATGATGATGATGATGATGATGATGATG 508
QY 996 ACATGAACCTCAACAGAACCTCG---GACTTTCGCTCAATTTCTATGAGCTGACT 1052
DB 509 ACCAGTACATTTGAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 568
QY 1053 GTGTGAGCAGAGTGGCAAGACATTCGATGATGATGATGATGATGATGATGATGATG 1112
DB 569 GATTCAGTGTGAGCAACAGAGACATGATGATGATGATGATGATGATGATGATGATG 628
QY 1113 CGGTAAATATGATTAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1172
DB 629 GTCTTCTGTGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 688
QY 1173 AGAAGAGGAGAGAGAGCTCTCTCCACATTTTAAAGACTGATCTTCAAGACATCC 1232
DB 689 ATTAAGAGAGAGGTTTAAAGATTTGCCACCTTAAAGATATGATGATGATGATGATG 748
QY 1233 CTGATGCTCTTTTGTGATGCTGACATGATGATGATGATGATGATGATGATGATG 1292

DB 749 AGAAGATATATTTGTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 805
QY 1293 ACTGTTTGTGTGACAGAGCTCAAGATTAATGATTAACAGCTCTGTGATGATCAATCA 1352
DB 806 ATGTGAGGTTTCAAGTACAGCTCAAGATTAATGATTAACAGCTCTGTGATGATCAAG 865
QY 1353 ATATGATC 1361
DB 866 ACATCATTC 874

RESULT 9

US-10-012-055-1
; Sequence 1, Application US/10012055
; Patent No. US20020164750A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: 56634, A NOVEL HUMAN PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE
; FILE REFERENCE: 10448-114001
; CURRENT APPLICATION NUMBER: US/10/012,055
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 60/248,325
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3224
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73)...(1335)
US-10-012-055-1

Query Match 2.0%; Score 73.8; DB 9; Length 3224;
Best Local Similarity 48.8%; Pred. No. 9.3e-10;
Matches 297; Conservative 0; Mismatches 297; Indels 15; Gaps 3;

QY 756 ATCACTACATGACTTTCGTTCAAGACCTATGACCTGTGCTTCCGCTACTTCGGG 815
DB 350 ATCTGCCAGTCAATTCATGATCAAGAGATATGTCGCCAGTCTTCAGGAACTCCGTG 409
QY 816 AGCTATTGGTATCGGCCGATGATTAATCTGATTCCTCTGACAGAGCCCTGATTTG 875
DB 410 ATGATTTGGCATGATGACCAAGATTAATCTGATGCTTACCCGAAACCCGCCAGCG 469
QY 876 AACTCTGATGCTGTGAGCTAGTGTCTCTATTTATGATGTCGAGGAGCATGATTC 935
DB 470 AAGG-----TGAAGGAGTGAATGATGCTCTCTTATCTCTACGATCGGACTCTG 520
QY 936 TTATTAAGACATGCAACATTAAGAGCGGAATTTCTGCAAGAGCTGCTTCCAGATCT 995
DB 521 TCATCAAGAGATGATCAGAGGACATTCGATGATGATGATGATGATGATGATGATGATG 580
QY 996 ACATGAACCTCAACAGAACCTCG---GACTTTCGCTCAATTTCTATGAGCTGACT 1052
DB 581 ACCAGTACATTTGAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 640
QY 1053 GTGTGAGCAGAGTGGCAAGACATTCGATGATGATGATGATGATGATGATGATGATG 1112
DB 641 GATTCAGTGTGACAGAGACATGATGATGATGATGATGATGATGATGATGATGATGATG 700
QY 1113 CGGTAAATATGATTAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1172
DB 701 GTCTTCTGTGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 760
QY 1173 AGAAGAGGAGAGAGAGCTCTCTCCACATTTTAAAGACTGATCTTCAAGACATCC 1232
DB 761 ATTAAGAGAGAGGTTTAAAGATTTGCCACCTTAAAGATATGATGATGATGATGATG 820
QY 1233 CTGATGCTCTTTTGTGATGCTGACATGATGATGATGATGATGATGATGATGATG 1292

Db	124	AGGATATGAGCTTTCACAAAGAACACAGAAAGTATATATTTGGTGAAGAGGAGAAAAA	183
Qy	1266	ACGCTCTCTGTAAAGACCCCTGACCGTGTACTTTGGTGTGCAGAGCTTCMAAGATATAGC	1325
Db	184	TATTTCTGG---AGAACTGTGMAAGAGATGTGAGTTCTGTATGACGCTGAGATCATGG	240
Qy	1326	ATTACAGCCTCTTGATGTATCATCTCAATATATATAGATC	1361
Db	241	ACTACAGCCTTCTGCTAGCATCATCAGACATCATTC	276

RESULT 13
US-09-783-590-2746
Sequence 2746, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16,201
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420, 856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346, 731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2746
LENGTH: 345
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (43)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (48)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (49)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (62)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (84)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (153)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (158)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (163)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (171)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (174)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature

```

1 LOCATION: (204)
2 OTHER INFORMATION: n equals a,t,g, or c
3 NAME/KEY: misc feature
4 LOCATION: (218)
5 OTHER INFORMATION: n equals a,t,g, or c
6 NAME/KEY: misc feature
7 LOCATION: (265)
8 OTHER INFORMATION: n equals a,t,g, or c
9 NAME/KEY: misc feature
10 LOCATION: (272)
11 OTHER INFORMATION: n equals a,t,g, or c
12 NAME/KEY: misc feature
13 LOCATION: (278)
14 OTHER INFORMATION: n equals a,t,g, or c
15 NAME/KEY: misc feature
16 LOCATION: (318)
17 OTHER INFORMATION: n equals a,t,g, or c
18 NAME/KEY: misc feature
19 LOCATION: (326)
20 OTHER INFORMATION: n equals a,t,g, or c
21 NAME/KEY: misc feature
22 LOCATION: (328)
23 OTHER INFORMATION: n equals a,t,g, or c
24 NAME/KEY: misc feature
25 LOCATION: (331)
26 OTHER INFORMATION: n equals a,t,g, or c
27 JS-09-783-590-2746

```

```

Query Match Similarity      1.1%; Score 42.4; DB 10; Length 345;
Best Local Similarity      55.7%; Pred. No. 0.097;
Matches 107; Conservative 0; Mismatches 83; Indels 2; Gaps 2

QY 1322 ATGATTACAGCCTCTTGATGTCAATCCATAATATAGATCATGACACAGAAACCCCTTA 1381
DB 1 ATGGGCTACAGGCTGTGCTGTGGCGGTCGCAACATGACCACNAGGNGGANNNG-6A 59
QY 1382 AGCAGTGAACACAGTACATCAGTGTGATCTCGAAGACCGGCCCCCAAAAGCCTCTGTAT 1441
DB 60 GGNGCAGGGCCCCGAGAGCACCTTCANNTAGAAAGCGGCTGTGGGACGAAAGGCGCTTAC 119
QY 1442 TCCACAGCCATGATTCATCCAGGAGAGG-CTGACGGGGTGTACATGAGACTGA 1500
DB 120 TCCACGCGCATGAGTCCATCCAGGGTGGCGCCNCGAGNGGGGAGGCCATGANTGGA 179
QY 1501 TGACCATATGGG 1512
DB 180 TGACACGATGGG 191

RESULT 14
US-09-816-685-3
; Sequence 3, Application US/0981685
; Patent No. US20020053091A1
; GENERAL INFORMATION:
; APPLICANT: CHANDRAMOULISWARAN, Ishwar et al
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; FILE REFERENCE: CL000856
; CURRENT APPLICATION NUMBER: US/09/816,685
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 41104
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(41104)
; OTHER INFORMATION: n = A,T,C or G
US-09-816-685-3

```

Query Match 1.1%; Score 41.2; DB 10; Length 41104;
Best Local Similarity 57.9%; Pred. No. 4.8;
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 3579 TTAGGGAAGGAGGAGGACATTTTATTAGCTGTACTATTTGCCCTACTTTG 3638

DB 30381 TTAGAAATATGATGATACATTTCACACATTATATAGTCTTCTTTTATTTT 30440

OY 3639 TATGTCGAAATGCAATCAATATTAAGTCATATATGTTTAACTAATTAAC 3698

DB 30441 TATAGTTACTGCTGTGCTCAAAAATGACATTCACGTGATGATGATTCATTTCATT 30500

OY 3699 TTAATG 3704

DB 30501 TTGATG 30506

RESULT 15

US-09-770-791-692/c

; Sequence 692, Application US/09770791

; Patent No. US20020062014A1

; GENERAL INFORMATION:

; APPLICANT: Gorlach, Jorn

; APPLICANT: An, Yong-Qiang

; APPLICANT: Hamilton, Carol M.

; APPLICANT: Price, Jennifer L.

; APPLICANT: Raines, Tracy M.

; APPLICANT: Yu, Yang

; APPLICANT: Rameaka, Joshua G.

; APPLICANT: Page, Amy

; APPLICANT: Mathew, Abraham V.

; APPLICANT: Ledford, Brooke L.

; APPLICANT: Moessner, Jeffrey P.

; APPLICANT: Haas, William David

; APPLICANT: Garcia, Carlos A.

; APPLICANT: Krieker, Maja

; APPLICANT: Slader, Ted

; APPLICANT: Davis, Keith R.

; APPLICANT: Allen, Keith

; APPLICANT: Hoffman, Neil

; APPLICANT: Hurban, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; FILE REFERENCE: 2029 (PARA-018PRV)

; CURRENT APPLICATION NUMBER: US/09/770,791

; CURRENT FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: 60/178,480

; PRIOR FILING DATE: 2000-01-27

; NUMBER OF SEQ ID NOS: 999

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 692

; LENGTH: 351

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-770-791-692

Query Match 1.1%; Score 39.4; DB 10; Length 351;

Best Local Similarity 57.9%; Pred. No. 0.66;

Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 1564 CATCATGACATCTACACTCTTACAGTTTGTAGAAAGTTGAGACACTTGGAAAGC 1623

DB 351 CATCATGACATCTCTCAAGATGACATGAGAAAGTGAACATACCTTGCAAATC 292

OY 1624 CTTGTCATGACGAGACACTGCTCAGTCATGCGCCAGGCTTCTACGCTGAACGGTT 1683

DB 291 TATGAAATATGATCCAAATGACATTCACGCAATCGAGCCTACACTTACTTAACGATT 232

OY 1684 C 1684

DB 231 C 231

Search completed: January 8, 2003, 18:13:03
Job time : 260 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 12:58:05 ; Search time 4885 Seconds
(without alignments)
12309.896 Million cell updates/sec

Title: US-10-003-354-3

Perfect score: 3713
Sequence: 1 attacagcgctggttagg.....aaacttaatgagttattta 3713

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hlc: *
9: gb_estl: *
10: gb_estl2: *
11: gb_hlc: *
12: gb_estl3: *
13: gb_estl4: *
14: gb_estl5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vit: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rtd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	859.4	23.1	889	14	BQ894466
2	841	22.7	1045	13	BM562816
3	828.2	22.3	1119	13	BM545034
4	827	22.3	865	9	AL522833
5	820.2	22.1	838	9	AU132382
6	817	22.0	1069	13	BM549810

7	813.6	21.9	824	9	AU126052	AU126052
8	802	21.6	814	9	AL528200	AL528200
9	800.6	21.6	904	14	BQ439604	BQ439604
10	797.6	21.5	1085	13	BM562145	BM562145
11	779.2	21.0	1046	13	BM455340	BM455340
12	772.8	20.8	864	13	BI090667	BI090667
13	768.8	20.7	824	9	AU130897	AU130897
14	766	20.6	908	14	BQ930009	BQ930009
15	760.4	20.5	783	9	AL522834	AL522834
16	754	20.3	876	14	BQ962454	BQ962454
17	721.4	19.4	882	9	AU149586	AU149586
18	715.8	19.3	875	13	BI181582	BI181582
19	694.2	18.7	919	14	BQ423398	BQ423398
20	687.2	18.5	2105	9	AL514476	AL514476
21	683.6	18.4	2105	11	BC020365	BC020365
22	680.4	18.3	2472	11	AK014403	AK014403
23	674	18.2	964	12	BE794576	BE794576
24	671.4	18.1	974	12	BG117080	BG117080
25	670.6	18.1	808	12	BG724319	BG724319
26	670.2	18.1	1043	13	BM461538	BM461538
27	665.4	17.9	705	9	AU130473	AU130473
28	660.4	17.8	803	12	BG679864	BG679864
29	654.4	17.6	683	13	BI1862007	BI1862007
30	653.4	17.6	685	9	AL710849	AL710849
31	651.6	17.5	736	14	BM982986	BM982986
32	649.4	17.5	768	12	BG717502	BG717502
33	648.4	17.5	682	14	BQ448271	BQ448271
34	643.8	17.3	1182	12	BG119715	BG119715
35	638.8	17.2	737	12	BG392973	BG392973
36	629.6	17.0	671	13	BI259656	BI259656
37	616	16.6	683	13	BM015907	BM015907
38	615.2	16.6	973	12	BG025366	BG025366
39	614.8	16.6	1065	14	BQ242439	BQ242439
40	609.2	16.4	629	14	BQ447995	BQ447995
41	609.2	16.4	1108	13	BI117425	BI117425
42	607.6	16.4	780	10	BE535818	BE535818
43	604	16.3	734	10	AM364156	AM364156
44	598.2	16.1	871	12	BF036836	BF036836
45	594	16.0	646	9	AL134930	AL134930

ALIGNMENTS

RESULT 1
LOCUS BQ894466
DEFINITION AGENCOURT_8727378 NIH_MGC_47 Homo sapiens CDNA clone IMAGE:6340661
5', mRNA sequence.
ACCESSION BQ894466
VERSION BQ894466.1 GI:22286480
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 889)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.llnl.gov
Plate: LCM2537 row: h column: 06
High quality sequence stop: 628.
Location/Qualifiers
1. 889


```

|||||
Db 181 CTGCTACATGACGAGACACTGCTCATGCTGCCAGGCTTCTACGCTGAAGGCTTC 240
Oy 1885 CAGGCGTTTCATGTCACACAGATATTTAGAAAGTCCCTTGAAGCCCTTCCCTCCAAA 1744
Db 241 CAGGCGTTTCATGTCACACAGATATTTAGAAAGTCCCTTGAAGCCCTTCCCTCCAAA 300
Oy 1745 AAGTTTGGTGGCTCATCTTTCTCTGGGAGAGAGGCTCCAGTGGCACTCTGCAAT 1804
Db 301 AAGTTTGGTGGCTCATCTTTCTCTGGGAGAGAGGCTCCAGTGGCACTCTGCAAT 360
Oy 1805 ACTTACACAGCCATGCTCTGCGGAGACACAAGGCAAGTACACAAGGAGAGAGT 1864
Db 361 ACTTACACAGCCATGCTCTGCGGAGACACAAGGCAAGTACACAAGGAGAGAGT 420
Oy 1865 GAGCGAGGCTTCACTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1924
Db 421 GAGCGAGGCTTCACTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 480
Oy 1925 ATCAGTGAAGGCTGCGCTATTTCTGACCCAGTTTCTGACCTCTAGTTGAGAGACTTTG 1984
Db 481 ATCAGTGAAGGCTGCGCTATTTCTGACCCAGTTTCTGACCTCTAGTTGAGAGACTTTG 540
Oy 1985 CAAATGCTTACTACAGATACAACTTGGAAAAGCTTGAAGTTGACAGTCAAGTTCACC 2044
Db 541 CAAATGCTTACTACAGATACAACTTGGAAAAGCTTGAAGTTGACAGTCAAGTTCACC 600
Oy 2045 CATTAAGGCGGAAAGCTAGAAAGACCTGGAACAAAGATCTGCACTCTGATCCCAAG 2104
Db 601 CATTAAGGCGGAAAGCTAGAAAGACCTGGAACAAAGATCTGCACTCTGATCCCAAG 660
Oy 2105 ATGTCAGGCGCTTGGCCCGACGATGCTGAATTTCTTCTACTTGGTCACTCAAAAAGAGT 2164
Db 661 ATGTCAGGCGCTTGGCCCGACGATGCTGAATTTCTTCTACTTGGTCACTCAAAAAGAGT 720
Oy 2165 GTAAATGAGTGAAGGAGGCTGCTCTCCATCTTCTCTGGAAGAAGCTTCTCTCT 2224
Db 721 GAAATGAGGAGGAGGAGGAGGCTGCTCTCCATCTCTCTGGAAGAAGCTTCTCTCT 780
Oy 2225 TCCCTTCCCTCATGATGATGGGCTTATGAGCTCAGAGATGAGGAGGAGGAGGAGGCT 2284
Db 781 TCCCTTCCCTCATGATGAGGCTTATGAGCTCAGAGATGAGGAGGAGGAGGAGGCT 840
Oy 2285 CACTCCAGAGTGGGGTGG--TACGGAATTTTCAACTGGCC---AACCTTGGCTCTCAT 2338
Db 841 CACTCCAGAGTGGGGTGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
Oy 2339 ATT-GAATTTTTCAGACCCCATTTCTTCATGCTGGAATGGGATGCTGGACTTGGCA 2397
Db 901 ATTGGAATTTTTCAGAAACCCCTTTCTTCATGATGGAATAAATGGGAATTTGGCT 960
Oy 2398 GCTTCTTTTCCCTGCTCT 2418
Db 961 GGGGACTTGGGAAAGTTTTT 981

RESULT 3
BM545034 1119 bp mRNA linear EST 20-FEB-2002
LOCUS BM545034
DEFINITION 5', mRNA sequence.
ACCESSION BM545034
VERSION BM545034.1 GI:18776787
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1119)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

```

Email: cgaabs-remail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LAM12360 row: 3 column: 01
High quality sequence start: 32.
High quality sequence stop: 652.
location/Qualifiers
1. 1119
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5588832"
/clone_11b="NIH-MGC_125"
/lab_host="DH10B"
/note="Organ: ovary (pool of 3); Vector: pCMV-Sport6;
Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool
of three ovaries, from females ranging in age from 38 to
49 yo. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 2.1 kb, insert size range 1-3.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 036."
BASE COUNT 276 a 239 c 246 g 357 t 1 others
ORIGIN
Query Match 22.3%; Score 828.2; DB 13; Length 1119;
Best Local Similarity 94.8%; Pred. No. 4; 7e-180;
Matches 890; Conservative 0; Mismatches 43; Indels 6; Gaps 3;
Oy 2780 CTGAAGGAATGAGAGAGTGGGACATGGGGTAATCTTATCCCTTTGTTAAAGAGAG 2839
Db 16 CGATGGAATGAGAGAGAGTGGGACATGGGGTAATCTTATCCCTTTGTTAAAGAGAG 75
Oy 2840 GCAGCAATGGGCTGGGAGATCATGCCCTTCTAGGCAATCTGTTACGCGCAGCT 2899
Db 76 GCAGCAATGGGCTGGGAGATCATGCCCTTCTAGGCAATCTGTTACGCGCAGCT 135
Oy 2900 ATAGTAATTTACTATTTTGCATTTGAAATATATCTGTTGTTTCTAAATGTGA 2959
Db 136 ATAGTAATTTACTATTTTGCATTTGAAATATATCTGTTGTTTCTAAATGTGA 195
Oy 2960 GACTTACCAATGAATTTAGATCATCTCCAGAGAGATTTTCTGCTCTTCATCT 3019
Db 196 GACTTACCAATGAATTTAGATCATCTCCAGAGAGATTTTCTGCTCTTCATCT 255
Oy 3020 TTTCACAGTGTCTCCGTTTGGAGCTAAGGTAAGAGGAGCACTTGTGCTGT 3079
Db 256 TTTCACAGTGTCTCCGTTTGGAGCTAAGGTAAGAGGAGCACTTGTGCTGT 315
Oy 3080 TTAACAGAGTCCATATCTGTGAGGCAAGCAATATTTTCTTAACTCATGGGAGACAG 3139
Db 316 TTAACAGAGTCCATATCTGTGAGGCAAGCAATATTTTCTTAACTCATGGGAGACAG 375
Oy 3140 CAGATTTCTGCTTGTGAGTGCATTTGCTGCTCATATGCTTACCCCTCTTTCATG 3199
Db 376 CAGATTTCTGCTTGTGAGTGCATTTGCTGCTCATATGCTTACCCCTCTTTCATG 435
Oy 3200 CAGGGAAGTGGAAATGGGGGCTAATATGCGCTCCCTCCCGCTCAAGAGTTGG 3259
Db 436 CAGGGAAGTGGAAATGGGGGCTAATATGCGCTCCCTCCCGCTCAAGAGTTGG 495
Oy 3260 TTTTCCATCTGATCCCTTCCACCTTGTGAGGAGAAAGGGGCTGATCTCAGGAG 3319
Db 496 TTTTCCATCTGATCCCTTCCACCTTGTGAGGAGAAAGGGGCTGATCTCAGGAG 555
Oy 3320 ATTGTTGAATTTCTGTTCTATCTCTATCCCAACCTGCTTGATATATGTTAGCC 3379
Db 556 ATTGTTGAATTTCTGTTCTATCTCTATCCCAACCTGCTTGATATATGTTAGCC 615

```

OY	3380	CATACC	CCCAAAATTAATGCTCTTATATTAAGACACCCAGCAGTTTGTGGCTGCTCTTT	3439
Db	616	CATACC	CCCAAAATTAATGCTCTATATTAGACACCCAGCAGTTTGTGGCTGCTCTTT	675
OY	3440	GCTGCC	ATGTTTTTACAGAAGGAAGTAATCTTGCTATTTTTTTTTTCATATTTACTA	3499
Db	676	GCTGCC	AGTTTTTTACAGAAGGAAGTAATCTTGCTATTTTTTTTTTCATATTTACTA	735
OY	3500	TTTATG	ATGATTTTAAAGTTTTTTATTAAGACAGAGTCTGTAGGGTGGAGGGAATA	3559
Db	736	TTTATG	ATGATTTTAAAGTTTTTTATTAAGACAGAGTCTGTAGGGTGGAGGGAATA	795
OY	3560	TTT---	-GAGGAGGGCTGGGCTTTAGGGAAGAATGGGAAAGCAACATTTTATTAG	3615
Db	796	TTTGAAG	GGAGGGGCTGGGCTTTAGGGAAGAATGGGGAAGCAACATTTTATTAG	855
OY	3616	TGTTACT	ATTTWG-CCCTACTTTGTAATGTTCAG-AAATGGCAAAATACAAATAAAGTG	3673
Db	856	GGGTACT	ATTTWGCCCTCCACATTGGATGTTCAGAAATGGCAATTTACATTTAAAGGG	915
OY	3674	ATATATG	TTTTAATGTATAATAAAGTTTAAAGTATATT 3712	
Db	916	TTTAAATG	GTTTTTATATGAATAAATCACTTTAAAGTATATT 954	

RESULT 4	AL522833/c	LOCUS	DEFINITION
	AL522833	865 bp	mRNA
	AL522833	LTI_NFL004_NBC2	linear
		Homo sapiens CDNA clone CSDB009YB17 3	EST 13-FEB-2001
		prime, mRNA sequence.	

ACCESSION	AL522833	
VERSION	AL522833.1	GI:12786326
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 865)
TITLE	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
JOURNAL	Full-length cDNA libraries and normalization
COMMENT	Unpublished (2001)
	Contact: genoscope

Genoscope - Centre National de Séquençage
BP 191 91006 Evry cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 865.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB009YB17"
/clone_1lb="LRT_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

```

Query Match	22.3%;	Score 827;	DB 9;	Length 865;
Best Local Similarity	97.6%;	Pred. NO. 8.6e-180;		
Matches 844;	Conservative	4;	Mismatches 16;	Indels 1;
				Gaps 1
QY	2809	GTATCTTTATCCCTTTTGTATAAACAGCA-GGCACCCATGGGCTGGAGATCATGACC	2867	

[illegible]

RESULT 5	
AU132382	
LOCUS	838 bp mRNA linear EST 01-AUG-2002
DEFINITION	AU132382 NT2RP3 Homo sapiens cDNA clone NT2RP3004516 5', mRNA
ACCESSION	AU132382
VERSION	AU132382.1 GI:10992736
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 838)

AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.

TITLE HRI human cDNA project

JOURNAL Unpublished (2000)

COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp

FEATURES HRI human cDNA project; 5'-3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

source Location/Qualifiers
1. 838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3004316"
/clone_11b="NT2RP3"
/cell_type="teratocarcinoma"
/note="Vector: pME18SF13; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"

BASE COUNT 227 a 200 c 199 g 210 t 2 others

ORIGIN

Query Match 22.1%; Score 820.2; DB 9; Length 838;
Best Local Similarity 99.5%; Pred. No. 3.2e-178;
Matches 833; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Oy 827 ATCCGGCCCGAGTACTTGTATCCCTGCGAGTGAAGCCGATGTAAGTCTGTAGC 886
Db 1 ATCCGGCCCGAGTACTTGTATCCCTGCGAGTGAAGCCGATGTAAGTCTGTAGC 60

Oy 887 TCTGAGAGTAGTGTCCCTATTCATTCATTCAGCAGCAGATGATTCATTTAAGACA 946
Db 61 TCTGAGAGTAGTGTCCCTATTCATTCATTCAGCAGCAGATGATTCATTTAAGACA 120

Oy 947 GTCCAACTTAAGAGCGGAGTTTCTGCAAGAGTCTCCAGATACATGAACCTC 1006
Db 121 GTCCAACTTAAGAGCGGAGTTTCTGCAAGAGTCTCCAGATACATGAACCTC 180

Oy 1007 AACCGAACCCTCGAGTTCGCTTAAATCTATGAGCTACTGTGTGACGACAGT 1066
Db 181 AACCGAACCCTCGAGTTCGCTTAAATCTATGAGCTACTGTGTGACGACAGT 240

Oy 1067 GGCAGAAACATTCGATTTGTGTGATGAACAATCTTTACCAAGATCGGTAAATGCA 1126
Db 241 GGCAGAAACATTCGATTTGTGTGATGAACAATCTTTACCAAGATCGGTAAATGCA 300

Oy 1127 ATCAAAATGACCTCAAGAGCTCAACCTACAAAGCGCGGCTTCCAGAAAGCGAGAG 1186
Db 301 ATCAAAATGACCTCAAGAGCTCAACCTACAAAGCGCGGCTTCCAGAAAGCGAGAG 360

Oy 1187 AAGCCTCTCCACATTTAAAGACCTAGACTCTTCAAGACATCCGTGATGCTTTT 1246
Db 361 AAGCCTCTCCACATTTAAAGACCTAGACTCTTCAAGACATCCGTGATGCTTTT 420

Oy 1247 TTGATGCTGACATGTACAGAGCTCTCTGTAGAGACCTGACAGCTGAGTTGGTGCG 1306
Db 421 TTGATGCTGACATGTACAGAGCTCTCTGTAGAGACCTGACAGCTGAGTTGGTGCG 480

Oy 1307 CAGAGCTCAAGATTAATGATTTACAGCCTCTGATGTCATTCATTAATATGATATGCA 1366
Db 481 CAGAGCTCAAGATTAATGATTTACAGCCTCTGATGTCATTCATTAATATGATATGCA 540

Oy 1367 CAACGAGAGCCCTTAAGCAGTGAACACAGTACTAGTTGATGATCGAAGACGGGCCCC 1426
Db 541 CAACGAGAGCCCTTAAGCAGTGAACACAGTACTAGTTGATGATCGAAGACGGGCCCC 600

Oy 1427 CAAAAGCTCTGTATTCACAGCCATGGAATCCATCCAGAGAGGCTTCGACGGGTGCT 1486
Db 601 CAAAAGCTCTGTATTCACAGCCATGGAATCCATCCAGAGAGGCTTCGACGGGTGCT 660

Oy 1487 ACCATGAGACATGATGACCATATGATGGTGCATCCCTGCCGGAATTAAGGGAAGG 1546
Db 661 ACCATGAGACATGATGACCATATGATGGTGCATCCCTGCCGGAATTAAGGGAAGG 720

Oy 1547 CTTTCGCTTTATATTCGATCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1606
Db 721 CTTTCGCTTTATATTCGATCATTCATTCATTCATTCATTCATTCATTCATTCATTC 780

Oy 1607 GAGCACTCTTGAAGAGCCCTGATACATGACGAGACACTGTCTCAGTGCATGCCCA 1663
Db 781 GAGCACTCTTGAAGAGCCCTGATACATGACGAGAC -CTGGCTCAGTGCATGCCCA 836

RESULT 6
BM549810 1069 bp mRNA linear EST 20-PEB-2002
LOCUS AGENCOURT 6543938 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5745824
DEFINITION 5', mRNA sequence.
ACCESSION BM549810
VERSION BM549810.1 GI:18785525
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1069)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM12769 row: 9 column: 09
High quality sequence start: 19
High quality sequence stop: 574.
Location/Qualifiers
1. 1069
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5745824"
/clone_11b="NIH_MGC_118"
/tissue_type="leukocyte"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."

BASE COUNT 263 a 290 c 216 g 298 t 2 others

ORIGIN

Query Match 22.0%; Score 817; DB 13; Length 1069;
Best Local Similarity 96.8%; Pred. No. 1.8e-177;
Matches 865; Conservative 0; Mismatches 26; Indels 3; Gaps 3;

Oy 1725 TGAAGCTTCTCTCC -AAAAGTTTCGGTCTGCTCATCTTCTCGCGGAGAGAGCT 1784
Db 35 TGAAGCTTCTCTCTCC -AAAAGTTTCGGTCTGCTCATCTTCTCGCGGAGAGAGCT 93

```

OY 1785 CCAGTGGCACTCTGCTGATTACTTACCAAGCCATCGTCTCTGGGGAACACAAAGCACAG 1844
DB 94 CCAAGTGGCACTCTGCTGATTACTTACCAAGCCATCGTCTCTGGGGAACACAAAGCACAG 153
OY 1845 TGACAAACAAAGGCAAGAGTGGAGCCAGCCCTTACCTTGGTCTGCTGATGTTTACCTC 1904
DB 154 TGACAAACAAAGGCAAGAGTGGAGCCAGCCCTTACCTTGGTCTGCTGATGTTTACCTC 213
OY 1905 AGACTCCAGCTTGGAGGAATGAGTGGAGCCAGCCCTTACCTTGGTCTGCTGATGTTTAC 1964
DB 214 AGACTCCAGCTTGGAGGAATGAGTGGAGCCAGCCCTTACCTTGGTCTGCTGATGTTTAC 273
OY 1965 CTCTAGTTGGAGAGACTTTCGAAATGCTAACACTACAGTACACCTTGGAAAGCTTGAAG 2024
DB 274 CTCTAGTTGGAGAGACTTTCGAAATGCTAACACTACAGTACACCTTGGAAAGCTTGAAG 333
OY 2025 TTGCGAGTCAAGATTAACCCATTAAAGCCGCAAAAGCCCTGCAAAAGCTGGAACAAGATTCT 2084
DB 334 TTGCGAGTCAAGATTAACCCATTAAAGCCGCAAAAGCCCTGCAAAAGCTGGAACAAGATTCT 393
OY 2085 GCCATCTCTGATGCCAAGATGTCAGCCCTTGGCCGAGCAATGCTGAATTTTCTTAC 2144
DB 394 GCCATCTCTGATGCCAAGATGTCAGCCCTTGGCCGAGCAATGCTGAATTTTCTTAC 453
OY 2145 TTGCTCATCAAAAAAGAGTGTAAATAGAGTGGAGGAGCTGCTCTCATCTTCTTCT 2204
DB 454 TTGCTCATCAAAAAAGAGTGTAAATAGAGTGGAGGAGCTGCTCTCATCTTCTTCT 513
OY 2205 GAAGAAGAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2264
DB 514 GAAGAAGAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 573
OY 2265 GAGGACCGCAGACATCCCTCCACATCCAGAGTGGGTGGTACGATTTTCAACTGGCCAAC 2324
DB 574 GAGGACCGCAGACATCCCTCCACATCCAGAGTGGGTGGTACGATTTTCAACTGGCCAAC 633
OY 2325 CCTTGGCTCCACATTTGAATTTTTCAGACCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 2384
DB 634 CCTTGGCTCCACATTTGAATTTTTCAGACCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 693
OY 2385 GCTGAGCTTGGAGAGCTTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2444
DB 694 GCTGAGCTTGGAGAGCTTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 753
OY 2445 TCAGACAGACTAGCTGGCAGATTAATCCCTACCTTACTCTCTCTCTCTCTCTCTCTCTCT 2504
DB 754 TCAGACAGACTAGCTGGCAGATTAATCCCTACCTTACTCTCTCTCTCTCTCTCTCTCTCT 813
OY 2505 GAATACCTCTGTAAATCTGTAAAGG-TTTTGGGGGATTAAGGGTTTAAACCACTGCC 2563
DB 814 -AATACCTCTGTAAATCTGTAAAGGTTTGGGGGATTAAGGGTTTAAACCACTGCC 872
OY 2564 AGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2617
DB 873 AACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 926

RESULT 7
LOCUS AUI26052 824 bp mRNA linear EST 01-AUG-2002
DEFINITION AUI26052 NT2RM4 Homo sapiens cDNA clone NT2RM4002631 5', mRNA
sequence
ACCESSION AUI26052.1 GI:10950768
VERSION AUI26052.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 824)
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.

```

```

TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
Y., Sugano,S., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project: 5'-3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source 1..824
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM4002631"
/clone_1lb="NT2RM4"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"
BASE COUNT 223 a 201 c 189 g 207 t 4 others
ORIGIN
Query Match 21.9%; Score 813.6; DB 9; Length 824;
Best Local Similarity 99.0%; Pred. No. 1,1e-176;
Matches 816; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
OY 721 CTTTCCAGTGAAGGGAGCACTGACCCCTGCTCATCTACATAGACTTTCGTTCAA 780
DB 1 CTTTCCAGTGAAGGGAGCACTGACCCCTGCTCATCTACATAGACTTTCGTTCAA 60
OY 781 GACCTATGACAGTGTGCTCCGCTACTCTCGGAGCTATTGTGATCGGCCGATGA 840
DB 61 GACCTATGACAGTGTGCTCCGCTACTCTCGGAGCTATTGTGATCGGCCGATGA 120
OY 841 TTACTGTATTCCTCTGACAGTGGCCGCTGATTAACCTCTGAGCTCTGAGCTAGTGG 900
DB 121 TTACTGTATTCCTCTGACAGTGGCCGCTGATTAACCTCTGAGCTCTGAGCTAGTGG 180
OY 901 TTCCCTATTCTATGTCGACGACCATGATCATTTAAGACAGTCCACATTAAGA 960
DB 181 TTCCCTATTCTATGTCGACGACCATGATCATTTAAGACAGTCCACATTAAGA 240
OY 961 GGGGAAATTCTGAGAAAGTGTTCAGAGATCTACATGAACTCAACAGAACCTCG 1020
DB 241 GGGGAAATTCTGAGAAAGTGTTCAGAGATCTACATGAACTCAACAGAACCTCG 300
OY 1021 GACTTTCCTGCTTAATTTCTATGAGCTGTACTGTGTGACGCGAGTGGCAAGCAATTGG 1080
DB 301 GACTTTCCTGCTTAATTTCTATGAGCTGTACTGTGTGACGCGAGTGGCAAGCAATTGG 360
OY 1081 GATGTGGTATGTAACAATCTTTTACCAAGATGGTAAATGCAATCAATATGAGCT 1140
DB 361 GATGTGGTATGTAACAATCTTTTACCAAGATGGTAAATGCAATCAATATGAGCT 420
OY 1141 CAAAGGCTCAACTTACAAGAGGGGGCTTCCAGAAAGAGCGAGAGAGCTTCTCCAC 1200
DB 421 CAAAGGCTCAACTTACAAGAGGGGGCTTCCAGAAAGAGCGAGAGAGCTTCTCCAC 480
OY 1201 AATTAAAGACTTACACTTCTTACAAGACATCCCTGATAGTCTTTTGGATGCTGACAT 1260
DB 481 AATTAAAGACTTACACTTCTTACAAGACATCCCTGATAGTCTTTTGGATGCTGACAT 540
OY 1261 GTACAAAGCTCTCTGTAAAGCTTGCAGAGCTGATGTTGGTGTGAGAGCTTCAAGAT 1320
DB 541 GTACAAAGCTCTCTGTAAAGCTTGCAGAGCTGATGTTGGTGTGAGAGCTTCAAGAT 600
OY 1321 AATGAGTTTACAGCTCTTGATGTCAATTCATTAATATGATCAGTACAGAGAGCCCTT 1380

```

```
|||||
Db 601 AATGGATTACAGCCTCTTGATGTCATCCATAATATAGATCATGCACAAAGAGAGCCCTT 660
|||||
QY 1381 AAGCAGTGAACACAGTACTAGTTGATCTCGAAGACCGCGCCCCCAAAAGGCTCTGTGA 1440
|||||
Db 661 AAGCAGTGAACACAGTACTAGTTGATCTCGAAGACCGCGCCCCCAAAAGGCTCTGTGA 720
|||||
QY 1441 TTCCACAGCCATCGAATCCATCCAGGAGAGGCTCGACGGGTGGTACCACTGAGACTGA 1500
|||||
Db 721 TTCCACAGCCATCGAATCCATCCAGGAGAGGCTCGACGGGTGGTACCACTGAGACTGA 780
|||||
QY 1501 TGACCATATGGGTGGCATCCCTCGCCCGAATAGTAAAGGGGAAA 1544
|||||
Db 781 TGACCATATGGGTGGCATNCTGGCCGAATAGTAAAGGGGAAA 824
|||||

RESULT 8
AL528200/c 814 bp mRNA linear EST 13-FEB-2001
LOCUS AL528200 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC025YF15 3
DEFINITION prime, mRNA sequence.
ACCESSION AL528200
VERSION AL528200.1 GI:12791693
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 814)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
1..814
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC025YF15"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : filiang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 259 a 183 c 164 g 195 t 13 others
ORIGIN
Query Match 21.6%; Score 802; DB 9; Length 814;
Best Local Similarity 98.0%; Pred. No. 5e-174;
Matches 798; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 2794 AGAGTGGGACATGGGTAATCTTATCCCTTTGTTAAACAGAGGAGCGCATGGGCTG 2853
|||||
Db 814 AGAGTGGGACATGGGTAATCTTATCCCTTTGTTAAACAGAGGAGCGCATGGGCTG 755
|||||
QY 2854 GGAGATCATAGCCCTTCCTAGGAGAGATCTGTTCAGTCCGAGGCTATAGTAATTATAC 2913
|||||
Db 754 GGAGATCATAGCCCTTCCTAGGAGAGATCTGTTCAGTCCGAGGCTATAGTAATTATAC 695
|||||
QY 2914 TATTTTGCATTTGAATATATTTCTGTTGTTTCTTAAATGTGAAGACTTACCAATGA 2973
|||||
Db 694 TATTTTGCATTTGAATATATTTCTGTTGTTTCTTAAATGTGAAGACTTACCAATGA 635
|||||
```

```
QY 2974 ATTTTATGATCATTTCTCCAGAGGAGATTTTTTTTCTCTCTCTCTCTCTCTCTCTCT 3033
|||||
Db 634 ATTTTATGATCATTTCTCCAGAGGAGATTTTTTTTCTCTCTCTCTCTCTCTCTCTCT 575
|||||
QY 3034 CTCCTGTTTGTGGAGCTTAAGGTAAGAGGAGGACACTTCTGCTGTTTAAACAGAGCTCCA 3093
|||||
Db 574 CTCCTGTTTGTGGAGCTTAAGGTAAGAGGAGGACACTTCTGCTGTTTAAACAGAGCTCCA 515
|||||
QY 3094 TATCTGTGAGGCCAGCAAAATATTTCTTAAACTCATGGGAGAGACAGAGATTCTTGCCTT 3153
|||||
Db 514 TATCTGTGAGGCCAGCAAAATATTTCTTAAACTCATGGGAGAGACAGAGATTCTTGCCTT 455
|||||
QY 3154 GGTGAGGTCAATGCTGTGCCATATGTCCTACCCCTGCTCTTCATGACGAGGAAGTTGGAA 3213
|||||
Db 454 GGTGAGGTCAATGCTGTGCCATATGTCCTACCCCTGCTCTTCATGACGAGGAAGTTGGAA 395
|||||
QY 3214 ATGGGGCTACATATGCCCTCTCTCCCTCTCAAGAGAGTTGGTTTCTCATCTGATC 3273
|||||
Db 394 ATGGGGCTACATATGCCCTCTCTCCCTCTCAAGAGAGTTGGTTTCTCATCTGATC 335
|||||
QY 3274 CTTCCTACTTGTTCAGGGGAGAGAGGGGCTGTATCTCAGGAGAGTTGTTGAATCCT 3333
|||||
Db 334 CTTCCTACTTGTTCAGGGGAGAGAGGGGCTGTATCTCAGGAGAGTTGTTGAATCCT 275
|||||
QY 3334 GTTCTATCCCTTCTCTATCCACCTTGCCTGATATAATATGTTAGCCCATACCCCAATAA 3393
|||||
Db 274 GATCTATCCCTTCTCTATCCACCTTGCCTGATATAATATGTTAGCCCATACCCCAATAA 215
|||||
QY 3394 CTGCTATATATAGACACACCCAGCAGTTCTGCTGCTGCTCTCTGCTGCTGCTGCTGCT 3453
|||||
Db 214 CTGCTATATATAGACACACCCAGCAGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 155
|||||
QY 3454 TACAAGAGGAAGAATTTCTGCTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3513
|||||
Db 154 TACAAGAGGAAGAATTTCTGCTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 95
|||||
QY 3514 AAGTCTTTTATTAAGGACAGAGTTCTGTTAGGGGTGGGAGGAATATTTAGGGAGGGCT 3573
|||||
Db 94 AAGTCTTTTATTAAGGACAGAGTTCTGTTAGGGGTGGGAGGAATATTTAGGGAGGGCT 35
|||||
QY 3574 GGGTCTTAGGGAAGGAATGGGAAGCAACATTT 3607
|||||
Db 34 GGGTCTTAGGGAAGGAATGGGAAGCAACATTT 1
|||||

RESULT 9
BQ439604
LOCUS BQ439604
DEFINITION 5', mRNA sequence.
ACCESSION BQ439604
VERSION BQ439604.1 GI:21178680
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 904)
NTH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov
Tissue Procurement: DCTD/DTG/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM13318 row: d column: 11
High quality sequence stop: 693.
```

Db	781	CAGCTCCTCTTTAAACTCCT 840
QY	2793	GAGAGTGGGACA---TGGGGTAATCTTTATCCCTTTTGTAAAA 2833
Db	841	NANAGTGGGGAACATGGNGGTAATCTTTATCCCTTTTGTATA 884
RESULT 10		
LOCUS	BM562145	1085 bp mRNA linear EST 20-FEB-2002
DEFINITION	AGENCOURT_6562112 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5745468	
ACCESSION	BM562145	
VERSION	BM562145.1	GI:18807983
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	1 (bases 1 to 1085)	
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM12768 row: h column: 13 High quality sequence stop: 601.	
FEATURES		
source	1..1085	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="IMAGE:5745468"
		/clone_lib="NIH_MGC_118"
		/tissue_type="leukocyte"
		/lab_host="DH10B"
		/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."
BASE COUNT	234 a 270 c 325 g 256 t	
ORIGIN		
Query Match 21.5%; Score 797.6; DB 13; Length 1085;		
Best Local Similarity 92.2%; Pred. No. 5,3e-173;		
Matches 963; Conservative 0; Mismatches 64; Indels 17; Gaps 11;		
QY	1	ATTAACAGCCCTGGTTAGGAAGGACGAGAGGGCGTTCGCTCCTTTGGGACTTTTCA 60
Db	40	ATTAACAGCCCTGGTTAGGAAGGACGAGAGGGCGTTCGCTCCTTTGGGACTTTTCA 99
QY	61	TGCCCTGTTTTTTTTCAGATGTGGCTTGGTCTGGCGCAAGTCCACGACGACGCTTA 120
Db	100	TGCCCTGTTTTTTTTCAGATGTGGCTTGGTCTGGCGCAAGTCCACGACGACGCTTA 159
QY	121	AGCTTACTCTCTGTAAGGGGAAGTATCCCTCTGGAAGCGGTTAAACTTGTGGAG 180
Db	160	AGCTTACTCTCTGTAAGGGGAAGTATCCCTCTGGAAGCGGTTAAACTTGTGGAG 219
QY	181	GGGCTGCGGAGCTGAGTTCTTCCCATCCAGGCGGAATGGTGTGGCCTTGAGCTGGTCC 240
Db	220	GGGCTGCGGAGCTGAGTTCTTCCCATCCAGGCGGAATGGTGTGGCCTTGAGCTGGTCC 279


```

|||||
Db 609 CACCTCATCAGCCTTGAAGAGTGCCATCAGTAGGCAATACCCACATGTGGGAGCC 668
QY 660 TGAGTACCAACAGAGCGTGATGCTCATGCAAGATTTCACGTGGTTGAGAGTATCT 719
|||||
Db 669 TGAGTACCAACAGAGCGTGATGCTCATGCAAGATTTCACGTGGTGAGAGTATCT 728
QY 720 TC-TTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCACTACAAATGACCTTCGTTTC 778
Db 729 TCTTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCACTACAAATGACCTTTCTTTC 788
QY 779 AGACCTATGCACCTGTTG-CCTTCGGCTACTTCCGGG 815
|||||
Db 789 AGACCTATGCACCTGTTGCCCTTCCGCTACTTCCGGG 826

RESULT 12
BI090667
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11027 row: i column: 06
High quality sequence spot: 793.
FEATURES
source
1. .864
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4998605"
/clone_lib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
BASE COUNT 214 a 196 c 184 g 269 t 1 others
ORIGIN
Query Match 20.8%; Score 772.8; DB 13; Length 864;
Best Local Similarity 97.2%; Pred. No. 2.7e-167;
Matches 829; Conservative 0; Mismatches 18; Indels 6; Gaps 4;

QY 2604 CACACGACACAAATTTCAAGCCATTTTCAGATCAGAACTCCAGAGTGTTCAGCAAGATG 2663
Db 1 CACACGACACAAATTTCAAGCCATTTTCAGATCAGAACTCCAGAGTGTTCAGCAAGATG 60
QY 2664 CCTATTCGTAGATTCCTCCAGAAAGCCATGTTTATGAGAGAGAGTAGTGATG 2723
Db 61 CCTATTCGTAGATTCCTCCAGAAAGCCATGTTTATGAGAGAGAGTAGTGATG 120
QY 2724 CTTCTGCCAGAACGAGCTCCTCTTTAACTCCCTCTCTTGTATGAATTTCTTAAGGCTGA 2783
Db 121 CTTCTGCCAGAACGAGCTCCTCTTTAACTCCCTCTCTTGTATGAATTTCTTAAGGCTGA 180

```

```

QY 2784 AGGAATGAAGAGAGTGGGACATGGGTAATCTTTATCCCTTTTGTAAACAGAGGCAG 2843
|||||
Db 181 AGGAATGAAGAGAGTGGGACATGGGTAATCTTTATCCCTTTTGTAAACAGAGGCAG 240
QY 2844 CATGGGCTGGGAGATCATAGCCCTTCCTAGGAGAAATCCTGTTCACGTCCAGGCCTATAG 2903
Db 241 CATGGGCTGGGAGATCATAGCCCTTCCTAGGAGAAATCCTGTTCACGTCCAGGCCTATAG 300
QY 2904 TAATTAATTAATTTTGAATTTGAATATATCTCGTGTGTTTCTTAATGTAAGACT 2963
Db 301 TAATTAATTAATTTTGAATTTGAATATATCTCGTGTGTTTCTTAATGTAAGACT 360
QY 2964 TACCAATGAATTTTAGATCATCTCTCAGAGGAGATTTTTCCTCTCTCTCTCTCTCTTTC 3023
Db 361 TACCAATGAATTTTAGATCATCTCTCAGAGGAGATTTCTTTTGTCTCTCTCTCTCTCTTTC 420
QY 3024 CACAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 3083
Db 421 CACAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 480
QY 3084 AGACAGTCCATATCTGTGAGGCGCAGCAATATTTTCTTAACTCATGGGAGACAGAGA 3143
Db 481 AGACAGTCCATATCTGTGAGGCGCAGCAATATTTTCTTAACTCATGGGAGACAGAGA 540
QY 3144 TTCTTGCCCTGTGAGGTCATCTGTGCAATATGTCCTACCCCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 3203
Db 541 TTCTTGCCCTGTGAGGTCATCTGTGCAATATGTCCTACCCCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 600
QY 3204 GAAGTTGGAATGGGGCTACATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 3262
Db 601 GAAGTTGGAATGGGGCTACATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 660
QY 3263 TCACATGTATCTTCCACCTTGTGAGGCGCAGCAATATTTTCTTAACTCATGGGAGACAGAGA 3322
Db 661 TCACATGTATCTTCCACCTTGTGAGGCGCAGCAATATTTTCTTAACTCATGGGAGACAGAGA 720
QY 3323 GTTGAATTCCTGTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 3382
Db 721 GTTGAATTCCTGTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTC 779
QY 3383 ACCCAATAA---CTGTCTATATAGACACCCCGCAGCAGTTTCGGCTGCTGCTCTTTC 3439
Db 780 ACCCAATAA---CTGTCTATATAGACACCCCGCAGCAGTTTCGGCTGCTGCTCTTTC 838
QY 3440 GCTGCCATGTTT 3452
Db 839 GCTGCCATGTTT 851

RESULT 13
AUI30897
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 824)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saiko,K., Kawai,Y.,
Yanamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975

```

Fax: 81-438-52-3986

Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES

Location/Qualifiers

source

1..824

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NT2RP3"

/cell_type="teratocarcinoma"

/cell_line="NT2"

/note="Vector: PME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"

BASE COUNT

227 a

190 c

209 t

4 others

Query Match

20.78; Score 768.8; DB 9; Length 824;

Best Local Similarity

99.08; Pred. No. 2.2e-166;

Matches 803; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

3;

Qy

834 CCGATGATTACTTCTATCCCTCTGCAGTGCAGCGCTGATTGAACCTCTGTAGCTCTGGAG 893

Db 1 CCGATGATTACTTCTATCCCTCTGCAGTGCAGCGCTGATTGAACCTCTGTAGCTCTGGAG 60

Qy 894 CTAGTGTTCCCTATTCTATGTGTCCAGCGACGATGATGATTCATTATTAAAGACAGTCCCAAC 953

Db 61 CTAGTGTTCCCTATTCTATGTGTCCAGCGACGATGATTCATTATTAAAGACAGTCCCAAC 120

Qy 954 ATAAAGAGCGGAATTTCTGCAGAAAGCTGCTCCAGGATGATCATGAACTCAACCGA 1013

Db 121 ATAAAGAGCGGAATTTCTGCAGAAAGCTGCTCCAGGATGATCATGAACTCAACCGA 180

Qy 1014 ACCCTCGSACTTTCGCTGCTAAATTCATGACTGTACTGTGTCGACGCGAGTGCACGA 1073

Db 181 ACCCTCGSACTTTCGCTGCTAAATTCATGACTGTACTGTGTCGACGCGAGTGCACGA 240

Qy 1074 ACATTCGGATTGTGGTGATGAACAACTCTTTACCAAGATCGGTAAAAATGATATCAAT 1133

Db 241 ACATTCGGATTGTGGTGATGAACAACTCTTTACCAAGATCGGTAAAAATGATATCAAT 300

Qy 1134 ATGACCTCAAGGCTCAACCTACAAAGCGGGGCTTCCGAAAGAGGAGGAGAGGCTC 1193

Db 301 ATGACCTCAAGGCTCAACCTACAAAGCGGGGCTTCCGAAAGAGGAGGAGAGGCTC 360

Qy 1194 TTCCACATTTAAAGACTAGACTTCTTACAAAGATCCCTGATGCTTTTTTTGGATG 1253

Db 361 TTCCACATTTAAAGACTAGACTTCTTACAAAGATCCCTGATGCTTTTTTTGGATG 420

Qy 1254 CTGACATGTACAACGCTCTCTGTAAAGCCCTGCAGCGTGACTGTTGGTGCTGCAGAGCT 1313

Db 421 CTGACATGTACAACGCTCTCTGTAAAGCCCTGCAGCGTGACTGTTGGTGCTGCAGAGCT 480

Qy 1314 TCAGATAATGGATATACAGCCTCTTGATGTCATTCATCCATATATAGATCATCCACAG 1373

Db 481 TCAGATAATGGATATACAGCCTCTTGATGTCATTCATCCATATATAGATCATCCACAG 540

Qy 1374 AGCCCTTAAGCAGTGAACACAGTACTCAGTTGATCTCGAAGACCGGCCCCCAAAAGG 1433

Db 541 AGCCCTTAAGCAGTGAACACAGTACTCAGTTGATCTCGAAGACCGGCCCCCAAAAGG 600

Qy 1434 CTCTGTATTCCACAGCCATGGAATCCATCCAGGAGAGGCTCGACGGGGTGGTACCATTGG 1493

Db 601 CTCTGTATTCCACAGCCATGGAATCCATCCAGGAGAGGCTCGACGGGGTGGTACCATTGG 660

Qy 1494 AGACTGATGACCATATGGGTGGCATCCCTGCCCGGAATAGTAAAGGGGAAAGGCTTCGC 1553

Db 661 AGACTGATGACCATATGGGTGGCATCCCTGCCCGGAATAGTAAAGGGGAAAGGCTTCGC 719

Qy 1554 TTTATATTGGCATATTGACATTCTAC-AGTCTTTACAGTTTGTAAAGATTGGAGCAC 1612

|||||

Db 720 TTTATATTGGCATATTGACATTCTACAGTCTTACAGTTTGTAAAGATTGGAGCAC 779

Qy 1613 TCTT-GGAAGCCCTGGTACATACGGAGAC 1642

|||||

Db 780 TCTTGGGAAGNCCTGGTNCATGACGGAGAC 810

RESULT 14

BQ930009

LOCUS

DEFINITION

AGENCOURT_8950942 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6473043

5', mRNA sequence.

ACCESSION

BQ930009

VERSION

BQ930009.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 908)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM14007 row: p column: 04

High quality sequence stop: 569.

FEATURES

Location/Qualifiers

1..908

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6473043"

/clone_lib="NIH_MGC_71"

/tissue_type="leiomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 2.1 kb.

BASE COUNT

198 a

222 c

269 g

219 t

ORIGIN

Query Match

20.68; Score 766; DB 14; Length 908;

Best Local Similarity

96.5%; Pred. No. 1e-165;

Matches 870; Conservative 0; Mismatches 20; Indels 12; Gaps 8;

Qy

1 ATTAACAGCCGCTGGTTAGGAAGACGAGAGGCGCTTCCTCTTTGGACCTTTTCA 60

Db

3 ATTAACAGCCGCTGGTTAGGAAGACGAGAGGCGCTTCCTCTTTGGACCTTTTCA 62

Qy

61 TGCCTCGTTTTTTTTCAGATGTGGCTTGGTGTGGCCCAAGGTCCTCCAGCAGCCAGCTTA 120

Db

63 TGCCTCGTTTTTTTTCAGATGTGGCTTGGTGTGGCCCAAGGTCCTCCAGCAGCCAGCTTA 122

Qy

121 AGCTTACTCTCTGTGAAGGGGAAAGTATCCCTGTGTGGAACGGGTTAACTTGTGGAG 180

Db

123 AGCTTACTCTCTGTGAAGGGGAAAGTATCCCTGTGTGGAACGGGTTAACTTGTGGAG 182

Qy

181 GGGTTCGGGACGTGAGTTCTCCCATGCCAGGCGAATGTGTGCGCTTGGAGCTGGTCC 240

Db

183 GGGTTCGGGACGTGAGTTCTCCCATGCCAGGCGAATGTGTGCGCTTGGAGCTGGTCC 242

Qy

241 AGGAGCCGCTCGAGTGTCTGAGGAGG-CGCCGAGGGGGGGGAGGTGCCCCACAGA 299

Db

243 AGGAGCCGCTCGAGTGTCTGAGGAGGCCCCCGGAGGGGGGGAGGTGCCCCACAGA 302

Qy

300 ACGCGGGTCTCTGTAAGAGACGTTGGGAAGATTCCGTAAGAGAGAGGAGAACCGGAT 359

Thu Jan 9 08:54:08 2003

us-10-003-354-3.rst

Page 13

Search completed: January 8, 2003, 18:06:40
Job time : 4911 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 12:54:40 : Search time 9375 Seconds
(without alignments)
11526.263 Million cell updates/sec

Title: US-10-003-354-3
Perfect score: 3713
Sequence: 1 attaacagcggtgtagg.....aaacttaagtgattatta 3713

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_mu.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3713	100.0	3713	9	HSU78575	U78575 Human 68 kD
2	3374.8	90.9	3637	9	BC007833	BC007833 Homo sapi
3	3038.8	81.8	104531	9	HS0303A1	AL096800 Human DNA
4	3037.6	81.8	4262	9	BC028580	BC028580 Homo sapi
5	3037.2	81.8	182895	2	AC011078	AC011078 Homo sapi
6	2081.2	56.1	112804	9	AC006476	AC006476 Homo sapi
7	2047.2	55.1	225432	9	AF027390	AF027390 Homo sapi
8	2014	54.2	2133	9	HSU78576	U78576 Human 68 kD
9	1852	49.9	2010	9	HSU78577	U78577 Human 68 kD
10	1835	49.4	180189	9	AL365510	AL365510 Human DNA
11	1654.8	44.6	154526	2	AL592111	AL592111 Homo sapi
12	1654.8	44.6	159148	9	AL592424	AL592424 Human DNA
13	1618.2	43.6	184090	2	AC009564	AC009564 Homo sapi
14	1543.4	41.6	2553	10	BC031774	BC031774 Mus muscu
15	1537.2	41.4	2567	9	AK098097	AK098097 Homo sapi
16	1522.6	41.0	2403	10	D86177	D86177 Mouse mRNA
17	1259.8	33.9	1709	9	BC007005	BC007005 Homo sapi
18	1143	30.8	1143	9	HSU78578	U78578 Human 68 kD
19	772.8	20.8	2723	10	BC003763	BC003763 Mus muscu
20	735.4	19.8	3360	5	AY062923	AY062923 Xenopus l
21	683.6	18.4	4226	10	BC019138	BC019138 Mus muscu
22	680.4	18.3	2225	10	AB006916	AB006916 Mus muscu
23	674.4	18.2	5047	9	AB011161	AB011161 Homo sapi
24	633.4	17.1	63747	2	AC015754	AC015754 Homo sapi
25	610.8	16.5	630	9	HUWY158E12	D86176 Mouse mRNA
26	588.8	15.9	2378	10	BC034864	BC034864 Mus muscu
27	588.8	15.9	2414	10	BC048695	BC048695 Mus muscu
28	583.2	15.7	1651	10	AF048695	AF048695 Sequence 67
29	581.6	15.7	1623	6	A59517	A59517 Sequence 9
30	581.6	15.7	2764	9	HSSTM7	X92493 H.sapiens m
31	581.6	15.7	2764	9	HSSTM7	AC015754 Homo sapi
32	565.4	15.2	3318	9	BC030587	BC030587 Homo sapi
33	489.8	13.2	63747	2	AC015754	AC015754 Homo sapi
34	463.6	12.5	204857	2	AC084272	AC084272 Mus muscu
35	463.6	12.5	224447	10	AC087062	AC087062 Mus muscu
36	408	11.0	1899	6	AX333345	AX333345 Sequence
37	408	11.0	1899	9	HSU78581	U78581 Human type
38	357.4	9.6	3402	3	AY069204	AY069204 Drosophil
39	346.6	9.3	2980	3	AY058737	AY058737 Drosophil
40	346.6	9.3	85862	3	AC004433	AC004433 Drosophil
41	346.6	9.3	163012	3	AC008289	AC008289 Drosophil
42	346.6	9.3	171972	3	AC012388	AC012388 Drosophil
43	346.6	9.3	304247	3	AE003452	AE003452 Drosophil
44	343.4	9.2	2922	3	AF071417	AF071417 Drosophil
45	336.2	9.1	13653	2	AC018218	AC018218 Drosophil

ALIGNMENTS

RESULT 1
HSU78575
LOCUS
DEFINITION
Human 68 kDa type I phosphatidylinositol-4-phosphate 5-kinase alpha
mRNA, clone PIP5K1a1, complete cds.
ACCESSION
U78575
VERSION
U78575.1
KEYWORDS
GI:1743870
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
1 (bases 1 to 3713)
Loijens, J.C. and Anderson, R.A.
TITLE
Type I phosphatidylinositol-4-phosphate 5-kinases are distinct

members of this novel lipid kinase family
J. Biol. Chem. 271 (51), 32937-32943 (1996)

JOURNAL MEDLINE 97115834
PUBMED 8955136

REFERENCE 2 (bases 1 to 3713)
Loijens, J.C. and Anderson, R.A.

AUTHORS Direct Submission
TITLE Submitted (18-NOV-1996) Pharmacology, University of Wisconsin -
Madison, 1300 University Ave., Madison, WI 53706, USA

JOURNAL Location/Qualifiers
FEATURES 1. .3713

source /organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="Pip5Ktal"

/cclone_lib="human fetal brain 5'-stretch cDNA library
(Clontech # HL1149x)"

401..2050
/EC_number="2.7.1.68"

/note="isoforms, possibly alternatively spliced, encoded
by GenBank Accession Numbers U78576 and U78577"

/codon_start=1
/product="68 kDa type I phosphatidylinositol-4-phosphate
5-kinase alpha"

/protein_id="AAC50910.1"

/db_xref="GI:1743871"

/translation="MASASSGSSVGFSDPAVPSCITLSSASGIKRPMASEVPIYAS
GPIKIKHSRVDSSGETTYKKTSSALKGAIQIHTVSLSTKPERDVLMDFFV
VESIFPSSGNLTPAHYNDFRVQYAPVAFRPRELFGIRPDYLYSLCSEPLIEL
CSSGASGLFVSSDEFLIKTVQKHEAEFLQKLLPGVYMLNPNRLLPKFYLAC
VOAGKNIIVVMNLLPRSKMHIKYLKGSYKRASQERERKPLPTFFDLDELQD
IPDGLPADMYNALCKTLQDCLVQFKIMDSILMSIHNDHAQREPLSETQYS
LTPRPAPOKALYAMESIOEARRGTMETDDHMGIPARNKSGERLLYIGLIDI
LQSYRFKLLSHSKWALVHDGTVSVHRPGYAEFRQFMCTVFKPLPKPSKKE
RSGSFRRAGSSNSCITYQPSVSGEHAQVTTKAEVPGVHLGRPDVLPQTPLEE
ISEGSPIDPSPFPLVGETLQMLTSTTLEKVAESEFTH"

polva_signal 3691..3696
polva_site 3713

BASE COUNT 919 a 865 c 882 g 1047 t

ORIGIN

Query Match 100.0%; Score 3713; DB 9; Length 3713;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3713; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTAACAGGCGGTGTTAGGAAGACGAGAGGGCGTTCGCTCCTTTGGGACTTTTCA 60
|||||
Db 1 ATTAACAGGCGGTGTTAGGAAGACGAGAGGGCGTTCGCTCCTTTGGGACTTTTCA 60
|||||

QY 61 TGCCTCGTTTTTTTTCAGATCTGGCTTGGTCTGGCGGCAAGGTCCACAGCCACTTA 120
|||||
Db 61 TGCCTCGTTTTTTTTCAGATCTGGCTTGGTCTGGCGGCAAGGTCCACAGCCACTTA 120
|||||

QY 121 AGCTTACTCTTCTGTAAGAGGGAAGATATCCCTCTGGAAGCGGTAAACTTGTGGAG 180
|||||
Db 121 AGCTTACTCTTCTGTAAGAGGGAAGATATCCCTCTGGAAGCGGTAAACTTGTGGAG 180
|||||

QY 181 GGGGTCTGGGACGTAGTCTTCCCATCCAGGCGAATGGTGTGGCCTTGAGCTGGTCC 240
|||||
Db 181 GGGGTCTGGGACGTAGTCTTCCCATCCAGGCGAATGGTGTGGCCTTGAGCTGGTCC 240
|||||

QY 241 AGGAGCGGCTGCAGCTGTCTAGGAGGCGCGGAGGGGGGGGGAGGTGGCCACAGAA 300
|||||
Db 241 AGGAGCGGCTGCAGCTGTCTAGGAGGCGCGGAGGGGGGGGGAGGTGGCCACAGAA 300
|||||

QY 301 CGCGGGTCTCTGAAGAGACGTGTGGGAAGATTCGATTCCGAGAAGGAAGAACCGGATT 360
|||||
Db 301 CGCGGGTCTCTGAAGAGACGTGTGGGAAGATTCGATTCCGAGAAGGAAGAACCGGATT 360
|||||

QY 361 GAAAGAGAGCCAGCGCTGAGGGGAGGGGGCTGCTAAGATGGCGCTCCCTCCCGG 420
|||||
Db 361 GAAAGAGAGCCAGCGCTGAGGGGAGGGGGCTGCTAAGATGGCGCTCCCTCCCGG 420
|||||

QY 421 GCCGTCTCTTCGGTTCGGTTTTTCATCCTTTGATCCCGGGTCCCTTCCTGTACCTTGTCT 480
|||||

Db 421 GCCGTCTCTTCGGTTCGGTTTTTCATCCTTTGATCCCGGGTCCCTTCCTGTACCTTGTCT 480
|||||
QY 481 CTCAGCATCTGGAATCAAGAGACCCATGGCATCTGAGGTGCCTTATGCTCTTGGCATGCC 540
|||||
Db 481 CTCAGCATCTGGAATCAAGAGACCCATGGCATCTGAGGTGCCTTATGCTCTTGGCATGCC 540
|||||
QY 541 CATCAAGAAAATAGGCCATAGAAAGTGTGATTCCTCAGGAGAGACAACATATAAAAAGAC 600
|||||
Db 541 CATCAAGAAAATAGGCCATAGAAAGTGTGATTCCTCAGGAGAGACAACATATAAAAAGAC 600
|||||
QY 601 AACCTCATCAGCTTGAAGAGTGCATCCAGTTCAGCATTTACCCACACTGTGGGAGCCT 560
|||||
Db 601 AACCTCATCAGCTTGAAGAGTGCATCCAGTTCAGCATTTACCCACACTGTGGGAGCCT 560
|||||
QY 661 GAGTACCAAAACAGAGCGGTGATGCTCTCATGCAAGATTTCTACGTGGTGTGAGAGTATCTT 720
|||||
Db 661 GAGTACCAAAACAGAGCGGTGATGCTCTCATGCAAGATTTCTACGTGGTGTGAGAGTATCTT 720
|||||
QY 721 CTTTCCCACTGAAGGAGCAACCTGACCCCTGCTCATCACTACAATGACTTTCTGTTTCAA 780
|||||
Db 721 CTTTCCCACTGAAGGAGCAACCTGACCCCTGCTCATCACTACAATGACTTTCTGTTTCAA 780
|||||
QY 781 GACCTATGCACCTGTTCCTTCGGTACTTCCGGGAGCTATTTGGTATCCGGCCGATGA 840
|||||
Db 781 GACCTATGCACCTGTTCCTTCGGTACTTCCGGGAGCTATTTGGTATCCGGCCGATGA 840
|||||
QY 841 TTACTTGTATTCCTCTGCGAGTGAGCGGTGATTAACATCTCTAGTCTGAGGAGTAGTGG 900
|||||
Db 841 TTACTTGTATTCCTCTGCGAGTGAGCGGTGATTAACATCTCTAGTCTGAGGAGTAGTGG 900
|||||
QY 901 TTTCCCTATTTCTATGTGTCCAGGAGCATGAGTTCATTTAAGACAGTCCCAACATAAAGA 960
|||||
Db 901 TTTCCCTATTTCTATGTGTCCAGGAGCATGAGTTCATTTAAGACAGTCCCAACATAAAGA 960
|||||
QY 961 GCGGGAATTTCTGCAGAGACTGCTTCCAGGATACATGAACTCAACCAAGCCCTCG 1020
|||||
Db 961 GCGGGAATTTCTGCAGAGACTGCTTCCAGGATACATGAACTCAACCAAGCCCTCG 1020
|||||
QY 1021 GACTTTGCTGCCTAAATTTCTATGAGTGTGTGTGAGGAGGTGGAAGAACATTCG 1080
|||||
Db 1021 GACTTTGCTGCCTAAATTTCTATGAGTGTGTGTGAGGAGGTGGAAGAACATTCG 1080
|||||
QY 1081 GATTGTGTTGATGACAAATCTTTTACCAGATCGGTAAATAATGCATATCAAAATATGACCT 1140
|||||
Db 1081 GATTGTGTTGATGACAAATCTTTTACCAGATCGGTAAATAATGCATATCAAAATATGACCT 1140
|||||
QY 1141 CAAAGGCTCAACCTACAAACGGGGCTTCCAGAAAGAGCGAGAGCCCTTCTTCCAC 1200
|||||
Db 1141 CAAAGGCTCAACCTACAAACGGGGCTTCCAGAAAGAGCGAGAGCCCTTCTTCCAC 1200
|||||
QY 1201 ATTTAAAGACCTTAGACTTCTTACAAAGACATCCCTGATGCTCTTTTGGATGCTGACAT 1260
|||||
Db 1201 ATTTAAAGACCTTAGACTTCTTACAAAGACATCCCTGATGCTCTTTTGGATGCTGACAT 1260
|||||
QY 1261 GTACACGCTCTCTGTAGAACCCCTGCAGCGTGACTGTTTGGTGTGACAGAGCTTCAAGAT 1320
|||||
Db 1261 GTACACGCTCTCTGTAGAACCCCTGCAGCGTGACTGTTTGGTGTGACAGAGCTTCAAGAT 1320
|||||
QY 1321 AATGGATTACAGCTCTTGCATCAATCAATATAGATCATGACACAGAGAGCCCTT 1380
|||||
Db 1321 AATGGATTACAGCTCTTGCATCAATCAATATAGATCATGACACAGAGAGCCCTT 1380
|||||
QY 1381 AAGCAGTGAACACACAGTACTCAGTTGATCTCGAAGACCGGCCCCCAAAAGGCTCTGTA 1440
|||||
Db 1381 AAGCAGTGAACACACAGTACTCAGTTGATCTCGAAGACCGGCCCCCAAAAGGCTCTGTA 1440
|||||
QY 1441 TTCCACAGCCATGGAATCCATCCAGGAGAGGCTGCACGGGGTGGTACCATGGAGACTGA 1500
|||||
Db 1441 TTCCACAGCCATGGAATCCATCCAGGAGAGGCTGCACGGGGTGGTACCATGGAGACTGA 1500
|||||
QY 1501 TCACCATATGCGTGGCATCCCTGCCGGAAATAGTAAGGGGAAAGGCTTCTGCTTATAT 1560
|||||

Db	1501	TGACCATTGGGTGGCATCCCTCGCCGGAAATAGTAAGAGGGGAAAGGCTTCTGCTTTATAT	1556
Qy	1561	TGGCATCAATTGACATTTACAGTCTTACAGGTTTGTTAAGAAGTTGGAGACACTCTTTGGAA	1620
Db	1561	TGGCATCAATTGACATTTACAGTCTTACAGGTTTGTTAAGAAGTTGGAGACACTCTTTGGAA	1620
Qy	1621	AGCCCTGGTACATGACGGAGACACTGTCTCAGTGCATCGCCACGGCTTCTACGCTGAACG	1680
Db	1621	AGCCCTGGTACATGACGGAGACACTGTCTCAGTGCATCGCCACGGCTTCTACGCTGAACG	1680
Qy	1681	GTTCCAGCGCTTCATGCTGCAACACACAGTATTTTAAGAAGATTCCTTTGAAGCCTTCTCCCTTC	1740
Db	1681	GTTCCAGCGCTTCATGCTGCAACACACAGTATTTTAAGAAGATTCCTTTGAAGCCTTCTCCCTTC	1740
Qy	1741	CAAAAAGTTTTCGCTCTGGCTCATCTTTCTCTCGCGAGCAGGCTCCAGTGGCAACTCCTG	1800
Db	1741	CAAAAAGTTTTCGCTCTGGCTCATCTTTCTCTCGCGAGCAGGCTCCAGTGGCAACTCCTG	1800
Qy	1801	CATTACTTACAGCCATCGGCTCTCTGGGGAACACAAAGGCACAGTGACAAAGGCGAGA	1860
Db	1801	CATTACTTACAGCCATCGGCTCTCTGGGGAACACAAAGGCACAGTGACAAAGGCGAGA	1860
Qy	1861	AGTGGAGCCAGGGGTTACCTTTGTCCTGATGTTTACCTCAGACTCCACCTTTGGA	1920
Db	1861	AGTGGAGCCAGGGGTTACCTTTGTCCTGATGTTTACCTCAGACTCCACCTTTGGA	1920
Qy	1921	GGAAATCAGTGAGGGCTCGCCTATTCCCTGACCCAGTTTCTCAGCTCTAGTTGGAGAGAC	1980
Db	1921	GGAAATCAGTGAGGGCTCGCCTATTCCCTGACCCAGTTTCTCAGCTCTAGTTGGAGAGAC	1980
Qy	1981	TTTGCAAAATGCTAACTTACAAGTACAACCTTGGAAGGCTTGAAGTTGCAGAGTCAGAGTT	2040
Db	1981	TTTGCAAAATGCTAACTTACAAGTACAACCTTGGAAGGCTTGAAGTTGCAGAGTCAGAGTT	2040
Qy	2041	CACCCATTAAGCGCAAGGCTCAGAAGACCTTGGACACAGATTTGCGCATCTCTGTGATCC	2100
Db	2041	CACCCATTAAGCGCAAGGCTCAGAAGACCTTGGACACAGATTTGCGCATCTCTGTGATCC	2100
Qy	2101	CAAGATGTCAGCCCTTGCCCCAGCAATGCTGAATTTTCTTCTACTTGGTCTCATCAAAAAG	2160
Db	2101	CAAGATGTCAGCCCTTGCCCCAGCAATGCTGAATTTTCTTCTACTTGGTCTCATCAAAAAG	2160
Qy	2161	GAGTGTAAATAGAGTAGGGGAGCTGCTCCATCTTCTTCTGAAAGAAACCTTCTCTC	2220
Db	2161	GAGTGTAAATAGAGTAGGGGAGCTGCTCCATCTTCTTCTGAAAGAAACCTTCTCTC	2220
Qy	2221	TCCCTTCCCTCTCTCATGAATGGGGCTTAGTGCCCTCAGAGAGTTGAGGACCGCAGCATCC	2280
Db	2221	TCCCTTCCCTCTCTCATGAATGGGGCTTAGTGCCCTCAGAGAGTTGAGGACCGCAGCATCC	2280
Qy	2281	CTTCCACTCCAGATTTGGGTGGTACGGATTTTCAACTTGGCCCAACCCCTTTGCCCTCCACTAT	2340
Db	2281	CTTCCACTCCAGATTTGGGTGGTACGGATTTTCAACTTGGCCCAACCCCTTTGCCCTCCACTAT	2340
Qy	2341	TGAATTTTTTTTCAGACCCCAATCTTTCATGCTGGAAATGGGATTTGCTGAGCTTTGGCAGCT	2400
Db	2341	TGAATTTTTTTTCAGACCCCAATCTTTCATGCTGGAAATGGGATTTGCTGAGCTTTGGCAGCT	2400
Qy	2401	TTCTTTTCCCTCGTCTTTGACTAGGAACCGGACCTTTAAATTTCTCTCAGGACACACTAGCT	2460
Db	2401	TTCTTTTCCCTCGTCTTTGACTAGGAACCGGACCTTTAAATTTCTCTCAGGACACACTAGCT	2460
Qy	2461	GGCACATTTACCTTACCTTAGTTCTTTCTCTGACTCTGACTCTGGAAGAATACTCTGCTGAATC	2520
Db	2461	GGCACATTTACCTTACCTTAGTTCTTTCTCTGACTCTGACTCTGGAAGAATACTCTGCTGAATC	2520
Qy	2521	TCTGTAAGGTTTTTTGGGGGATAAGGGTGTTTTAAACCACCTCCCGAGCTTTCTTCTCTTTT	2580
Db	2521	TCTGTAAGGTTTTTTGGGGGATAAGGGTGTTTTAAACCACCTCCCGAGCTTTCTTCTCTTTT	2580
Qy	2581	TTTTTTTCTGAAAAAGGAAAAAGCACAGCACAAATTTCAAGCCATTTTCAGATCAGA	2640
Db	2581	TTTTTTTCTGAAAAAGGAAAAAGCACAGCACAAATTTCAAGCCATTTTCAGATCAGA	2640

Db 1073 GGACTTGGCTGCCATAATCTATGGACTGTACTGTCTCAGGCGAGGTGGCAAGAACATTC 1132
Qy 1080 GGATTGGGTGATGAACAATCTTTTACAGATCGGTAAATAATGCATATCAAAATATGACC 1139
Db 1133 GGATTGGGTGATGAACAATCTTTTACAGATCGGTAAATAATGCATATCAAAATATGACC 1192
Qy 1140 TCAAGGCTCAACCTACAAACGGCGGCTTCCCAAGAGAGCGAGAGAGGCTCTTCCCA 1199
Db 1193 TCAAGGCTCAACCTACAAACGGCGGCTTCCCAAGAGAGCGAGAGAGGCTCTTCCCA 1252
Qy 1200 CATTTAAAGACCTAGACTTCTTACAAGACATCCCTGATGGTCTTTTGGATGCTGACA 1259
Db 1253 CATTTAAAGACCTAGACTTCTTACAAGACATCCCTGATGGTCTTTTGGATGCTGACA 1312
Qy 1260 TGTACAGGCTCTCTGTAAGACCTCGAGCGTGAAGTGGTGGTGGTGGTGGTGGTGGTGGT 1319
Db 1313 TGTACAGGCTCTCTGTAAGACCTCGAGCGTGAAGTGGTGGTGGTGGTGGTGGTGGTGGT 1372
Qy 1320 TAATGGATTAACAGGCTCTGATGATCAATCCATAATATAGATCATGCAACGAGAGCGCT 1379
Db 1373 TAATGGATTAAGGCTCTGATGATCAATCCATAATATAGATCATGCAACGAGAGCGCT 1432
Qy 1380 TAAGCAGTGAACACAGTACTCACTGATGATCGAAGACCGCGCCGCCAAAGGCTCTGT 1439
Db 1433 TAAGCAGTGAACACAGTACTCACTGATGATCGAAGACCGCGCCGCCAAAGGCTCTGT 1492
Qy 1440 ATTCACAGGCTGAATCCATCCAGGAGAGGCTCGAGCGGGTGGTACATGAGAGCTG 1499
Db 1493 ATTCACAGGCTGAATCCATCCAGGAGAGGCTCGAGCGGGTGGTACATGAGAGCTG 1552
Qy 1500 ATGACCATATGGGTGGATCCCTCCCGGATAGTAAAGGGAAGGCTCTGCTTTATA 1559
Db 1553 ATGACCATATGGGTGGATCCCTCCCGGATAGTAAAGGGAAGGCTCTGCTTTATA 1612
Qy 1560 TTGGCATCATGACATCTACAGTCTTACAGTGTGTAAAGTGTGGAGCAGCTTTGGA 1619
Db 1613 TTGGCATCATGACATCTACAGTCTTACAGTGTGTAAAGTGTGGAGCAGCTTTGGA 1672
Qy 1620 AAGCCCTGGTACATGAGGAGAGCTGTCTCAGTGCATCGCCGAGGCTTACGCTGAAC 1679
Db 1673 AAGCCCTGGTACATGAGGAGAGCTGTCTCAGTGCATCGCCGAGGCTTACGCTGAAC 1732
Qy 1680 GGTTCACGCGCTCATGTGCAACACAGTATTTAAGAGAGATTCCTTGAAGCCTTCTCCT 1739
Db 1733 GGTTCACGCGCTCATGTGCAACACAGTATTTAAGAGAGATTCCTTGAAGCCTTCTCCT 1777
Qy 1740 CCAAAAAGTTTGGTCTGGCTCATCTTTCTCTCGCGGAGCAGGCTCCAGTGGCAACTCT 1799
Db 1778 ----- 1777
Qy 1800 GCATTACTTACCAGCCATCGGCTCTCGGGGACACAAAGGCAACAGTGACAAAGGAG 1859
Db 1778 ----- 1777
Qy 1860 AAGTGGAGCCAGGCTTACCTTGGTCTGCTGATGTTTACCTCAGACTCCACCTTTGG 1919
Db 1778 -----GGTTCACCTTGGTCTGCTGATGTTTACCTCAGACTCCACCTTTGG 1825
Qy 1920 AGGAAATCAGTGGGCTCGCTATTCTGACCCCGAGTTTCTACCTCTAGTTGGAGAGA 1979
Db 1826 AGGAAATCAGTGGGCTCGCTATTCTGACCCCGAGTTTCTACCTCTAGTTGGAGAGA 1885
Qy 1980 CTTTGAATAGCTAACATACAGTACAACCTTTGAAAAGCTTGAAGTGGCAGAGTCAGAGT 2039
Db 1886 CTTTGAATAGCTAACATACAGTACAACCTTTGAAAAGCTTGAAGTGGCAGAGTCAGAGT 1945
Qy 2040 TCACCCATTAAAGCCAAAGCTCAGAGACCTGGAACAGATTCCTGCCATCTCTGTATC 2099
Db 1946 TCACCCATTAAAGCCAAAGCTCAGAGACCTGGAACAGATTCCTGCCATCTCTGTATC 2005
Qy 2100 CCAAGATGTGAGCCCTTGGCCCGAGCAATGCTGAATTTTCTTACTTGGTCAATCAAAAA 2159
Db 2006 CCAAGATGTGAGCCCTTGGCCCGAGCAATGCTGAATTTTCTTACTTGGTCAATCAAAAA 2065

Qy 2160 GGAGTGTAAATAGAAGTGAAGGGAGCTGCTCCTCCATCTTCTCCTGAAGAAAGAACTTCT 2219
Db 2066 GGAGTGTAAATAGAAGTGAAGGGAGCTGCTCCTCCATCTTCTTCTTGAAGAAAGAACTTCT 2125
Qy 2220 CTCCTTCTCTCTCTCATGAATGGCCCTTAGTGGCTCAGAGAGTTGAGGACCGAGCATC 2279
Db 2126 CTCCTTCTCTCTCTCATGAATGGCCCTTAGTGGCTCAGAGAGTTGAGGACCGAGCATC 2185
Qy 2280 CCTCCACTCCAGAGTTGGGTGGTACGATTTTCAACTGGCCAAACCTTTGGCTCCACTA 2339
Db 2186 CCTCCACTCCAGAGTTGGGTGGTACGATTTTCAACTGGCCAAACCTTTGGCTCCACTA 2245
Qy 2340 TTGAATTTTTTCAGACCCCATTTCTCATGCTGGAATGGGATTCGTGGAGCTTGGCAGC 2399
Db 2246 TTGAATTTTTTCAGACCCCATTTCTCATGCTGGAATGGGATTCGTGGAGCTTGGCAGC 2305
Qy 2400 TTTCTTTCCCTCTCTGTTGACTAGGAACCGAGCTCTTAAATTTCCCTCAGGACAGACTAGC 2459
Db 2306 TTTCTTTCCCTCTCTGTTGACTAGGAACCGAGCTCTTAAATTTCCCTCAGGACAGACTAGC 2365
Qy 2460 TGGCACATTTATCCCTACCTTAGTTCTCTCTGACTCCCTGGAGAAATACTCCTGTAAAT 2519
Db 2366 TGGCACATTTATCCCTACCTTAGTTCTCTCTGACTCCCTGGAGAAATACTCCTGTAAAT 2425
Qy 2520 CTCGTAAAGTTTTTGGGGGATGAAGGTGTTTAAACCACTCCAGCTTCTTCTCTTCTT 2579
Db 2426 CTCGTAAAGTTTTTGGGGGATGAAGGTGTTTAAACCACTCCAGCTTCTTCTCTTCTT 2485
Qy 2580 TTTTCTTCTGAAAAAGGAAAAAGCACACACACAAATTTCAAGCCATTTTCAGATCAG 2639
Db 2486 TTTTCTTCTGAAAAAGGAAAAAGCACACACACAAATTTCAAGCCATTTTCAGATCAG 2544
Qy 2640 AACTCCAGAGTGTGCAAGATGCTTATGCTAGAGTTCCTCAGAAAGAGCCATGGTGT 2699
Db 2545 AACTCCAGAGTGTGCAAGATGCTTATGCTAGAGTTCCTCAGAAAGAGCCATGGTGT 2604
Qy 2700 TTATGAGAGAGAGTGTAGTATGCTCTGCGGAGAGAGCTCCTCTTTAAACTCTCTCTC 2759
Db 2605 TTATGAGAGAGAGTGTAGTATGCTCTGCGGAGAGAGCTCCTCTTTAAACTCTCTCTC 2664
Qy 2760 TCTTGTATGAATTTCTTAAAGCTGAAGGAAATGAAGAGAGCTGGGACATGGGTAAATCTTAT 2819
Db 2665 TCTTGTATGAATTTCTTAAAGCTGAAGGAAATGAAGAGAGCTGGGACATGGGTAAATCTTAT 2724
Qy 2820 CCTTTTGTAAAAAGAGGAGCAGCCATGGGCTGGAGATCATAGCCCTTCTTAGGCAGA 2879
Db 2725 CCTTTTGTAAAAAGAGGAGCAGCCATGGGCTGGGAGATCATAGCCCTTCTTAGGCAGA 2784
Qy 2880 ATCTGTCTACTGCCAGGCTATAGTAATATTATTTTGAATTTGAAATATATTTCTG 2939
Db 2785 ATCTGTCTACTGCCAGGCTATAGTAATATTATTTTGAATTTGAAATATATTTCTG 2844
Qy 2940 GTTGTTTTCTTAAATGTGAACACTTACCAAAATGAATTTTACATCATCTCCAGAGAGAT 2999
Db 2845 GTTGTTTTCTTAAATGTGAACACTTACCAAAATGAATTTTACATCATCTCCAGAGAGAT 2904
Qy 3000 TTTTGTGTCTCTCTCATCTTTTCCAAAGTGTCTCTCTCTGTTTGGAGCTAAGGTAAG 3059
Db 2905 TTTTGTGTCTCTCTCATCTTTTCCAAAGTGTCTCTCTCTGTTTGGAGCTAAGGTAAG 2964
Qy 3060 AGGGAGACATCTCTGCTTTTAAACAGACATCCATATCTGTGAGGCCAGCAAAATTTTC 3119
Db 2965 AGGGAGACATCTCTGCTTTTAAACAGACATCCATATCTGTGAGGCCAGCAAAATTTTC 3024
Qy 3120 TTAACATCATGGGAGACAGAGATTTCTGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3179
Db 3025 TTAACATCATGGGAGACAGAGATTTCTTGGCTTGGTGGTGGTGGTGGTGGTGGTGGT 3084
Qy 3180 CCTACCCCTCTCTCTCATGAGGAGGAGCTTGGAAATGGGGCTACATATGCTCTCTCTC 3239
Db 3085 CCTACCCCTCTCTCTCATGAGGAGGAGCTTGGAAATGGGGCTACATATGCTCTCTCTC 3144

QY	3240	CCGCTCACAAGAGTGTGGTGTTCACATCATGATCCCTTCCACTCTTGTGACGGGAAGAAGG	3299
Db	3145	CCCGCTCACAAGAGTGTGGTGTTCACATCATGATCCCTTCCACTCTTGTGACGGGAAGAAGG	3204
QY	3300	GGGCGCTGATATCAGCAGATGTTGAATCCCTGTTCTATCCCTTCTCTATCCCAACCT	3359
Db	3205	GGGCGCTGATATCAGCAGATGTTGAATCCCTGTTCTATCCCTTCTCTATCCCAACCT	3264
QY	3360	GCCTTGATAATATGTTAGCCATACCCCAATAACTGCTCTATATTAGACACACCCCGACCCA	3419
Db	3265	GCCTTGATAATATGTTAGCCATACCCCAATAACTGCTCTATATTAGACACACCCCGACCCA	3324
QY	3420	GTTTCTGGCTGCCGTCTTGTGTCGCCAATGTTTTTACAAGAAGAAGAAATCTTGCTAT	3479
Db	3325	GTTTCTGGCTGCCGTCTTGTGTCGCCAATGTTTTTACAAGAAGAAGAAATCTTGCTAT	3384
QY	3480	TTTTTTTTTCATAATTTACTATTTATGATCTATTTAACTGTTTATTAAAGACACAGTTCT	3539
Db	3385	TTTTTTTTTCATAATTTACTATTTATGATCTATTTAACTGTTTATTAAAGACACAGTTCT	3444
QY	3540	GTTAGGGGTGGGAGGAATATTTGAGGAGCGGTGGGTCTTAGGGAAGAAGAAATGGGGAAG	3599
Db	3445	GTTAGGGGTGGGAGGAATATTTGAGGAGCGGTGGGTCTTAGGGAAGAAGAAATGGGGAAG	3504
QY	3600	CAACATTTTATTAAAGTGTCTACTATTTGCCCTCTACTTTGTTGTTTCAGAAATGGCAAT	3659
Db	3505	CAACATTTTATTAAAGTGTCTACTATTTGCCCTCTACTTTGTTGTTTCAGAAATGGCAAT	3564
QY	3660	ACAATATAAAGTCATATATCGTTTAACTGTAATAAATCTTAATGAGTATTTA 3713	
Db	3565	ACAATATAAAGTCATATATCGTTTAACTGTAATAAATCTTAATGAGTATTTA 3618	
RESULT 3			
LOCUS	HSJD303A1	104531 bp	DNA linear PRI 17-MAY-2000
DEFINITION	Human DNA sequence from clone RP1-303A1 on chromosome 6. Contains		
	an XCRDR (cox sackle virus and adenovirus receptor) pseudogene, a		
	1-phosphatidylinositol-4-phosphate kinase (PTPDINS(4)P-5-kinase, EC		
	2.7.1.68) pseudogene, the last exon of a putative new gene, ESTs,		
	STSs, GSS and a putative cpG island, complete sequence.		
ACCESSION	AL096800		
VERSION	AL096800.20	GI:7242638	
KEYWORDS	HTG; CpG island; CXADR.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 104531)		
AUTHORS	Collier,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire,		
	CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk		
	requests: clonesrequest@sanger.ac.uk		
COMMENT	On Mar 14, 2000 this sequence version replaced gi:7159759.		
	During sequence assembly data is compared from overlapping clones.		
	Where differences are found these are annotated as variations		
	together with a note of the overlapping clone name. Note that the		
	variation annotation may not be found in the sequence submission		
	corresponding to the overlapping clone, as we submit sequences with		
	only a small overlap as described above.		
	This sequence has been finished according to sequence map criteria		
	as follows. An attempt is made to resolve all sequencing problems,		
	such as compressions and repeats, but not necessarily within known		
	annotated human repeat sequence elements (e.g. Alu). Where the		
	sequence is ambiguous, there is an annotation using the 'unsure'		
	feature key.		
	The following abbreviations are used to associate primary accession		
	numbers given in the feature table with their source databases:		
	Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information		
	on the WORMPEP database can be found at		
	http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence		
	was generated from part of bacterial clone contigs of human		

```
repeat_region 19495..19531
/notes="L2 repeat: matches 2617. .2653 of consensus"
repeat_region 19573..20005
/notes="MTC repeat: matches 1. .402 of consensus"
repeat_region 20263..20419
/notes="MER46C repeat: matches 2. .162 of consensus"
repeat_region 21023..21333
/notes="THE1B repeat: matches 1. .307 of consensus"
repeat_region 21335..27440
/notes="LIPAS repeat: matches 11. .6145 of consensus"
repeat_region 27446..27515
/notes="THE1B repeat: matches 294. .364 of consensus"
misc_feature 27921..28345
/notes="match: GSS: Em:AQ057927"
misc_feature 27956..28359
/notes="match: GSS: Em:AQ057921"
repeat_region 28549..28855
/notes="AluX repeat: matches 1. .309 of consensus"
repeat_region 30105..30229
/notes="FLAM.C repeat: matches 1. .124 of consensus"
misc_feature complement(30634..31065)
misc_feature complement(30644..31063)
repeat_region 30858..31142
/notes="AluJo repeat: matches 1. .294 of consensus"
misc_feature 31104..31403
/notes="match: GSS: Em:B63554"
repeat_region 32509..32782
/notes="AluJo repeat: matches 1. .284 of consensus"
repeat_region 33014..33181
/notes="FRAM repeat: matches 4. .170 of consensus"
repeat_region 33973..34274
/notes="AluX repeat: matches 1. .300 of consensus"
repeat_region 34836..35132
/notes="AluX repeat: matches 2. .296 of consensus"
repeat_region 35146..35195
/notes="23 copies 2 mer gt 84% conserved"
repeat_region 35160..35195
/notes="9 copies 4 mer gtgt 100% conserved"
repeat_region 35207..35258
/notes="26 copies 2 mer at 71% conserved"
repeat_region 35807..36268
/notes="L1A9 repeat: matches 5793. .6268 of consensus"
repeat_region 36269..36578
/notes="AluSg repeat: matches 1. .310 of consensus"
repeat_region 36579..36615
/notes="L1A9 repeat: matches 6268. .6303 of consensus"
repeat_region 36963..37278
/notes="AluX repeat: matches 1. .312 of consensus"
repeat_region 37957..38250
/notes="AluSg repeat: matches 1. .304 of consensus"
misc_feature complement(38069..38560)
repeat_region 38844..39046
/notes="L2 repeat: matches 2542. .2746 of consensus"
repeat_region 39048..39292
/notes="MER73 repeat: matches 376. .633 of consensus"
repeat_region 39293..39604
/notes="AluX repeat: matches 1. .300 of consensus"
repeat_region 39605..39959
/notes="MER73 repeat: matches 65. .376 of consensus"
repeat_region 39960..39995
/notes="18 copies 2 mer gt 100% conserved"
repeat_region 39998..40274
/notes="AluY repeat: matches 1. .277 of consensus"
repeat_region 40276..40348
/notes="MER73 repeat: matches 1. .74 of consensus"
repeat_region 40640..40778
/notes="LIME2 repeat: matches 5045. .5178 of consensus"
repeat_region 40788..40924
/notes="L1A10 repeat: matches 6180. .6316 of consensus"
repeat_region 40964..41130
```

```
repeat_region 41167..41424
/notes="AluJb repeat: matches 128. .296 of consensus"
repeat_region 41441..41608
/notes="LIME2 repeat: matches 5189. .5449 of consensus"
repeat_region 41609..41892
/notes="L1A9 repeat: matches 5884. .6057 of consensus"
repeat_region 41893..42166
/notes="AluX repeat: matches 1. .293 of consensus"
repeat_region 42909..43196
/notes="L1A9 repeat: matches 6057. .6308 of consensus"
repeat_region 43299..43875
/notes="AluY repeat: matches 1. .287 of consensus"
repeat_region 44188..44521
/notes="LIME2 repeat: matches 5478. .6155 of consensus"
repeat_region 44580..44725
/notes="MUT1A repeat: matches 1. .363 of consensus"

Query Match 81.8%; Score 3038.8; DB 9; Length 104531;
Best Local Similarity 91.2%; Pred. No. 0;
Matches 3420; Conservative 0; Mismatches 277; Indels 55; Gaps 16;

QY 2 TTACAGGCGGTGGTTAGGAAGGACGAGAGGGCGGTTCGCTCTCTTTGGGACTTTTCAT 61
Db 64268 TAAACAGGCGGTGGTTAGGAGGACAGAGAGCGCGTTCGCTCTCTGGGACTTTTCAC 64327
QY 62 GCCTCG-----TTTTTTTTTCAGATGTGGCTTGGGCGCAAGTCCACAGCAGCC 114
Db 64328 GCCTCGTTTGTGTTTCTTCAGATGTGGCTTCTTCTGACGAGGGTCCAAGCAGTC 64387
QY 115 AGCTTAAGCTTACTTCTGTGTAAGGGGAAAGTATCCCTGTGTGGAAGCGTTAAACTT 174
Db 64388 AGCTTGAGCTTACTTCTGTGTAAGGGGAAAGTCTCTCTGAGGAAGCGGTTGAACCT 64447
QY 175 GTGAGAGGGGTGCGGAGCGTGTCTTCCCATCCAGGCGAATGTGTGGCTTGCAGC 234
Db 64448 GTGAGGCGGTGCGAGAGCGGCGTCTCCCATCCAGGCGAATGTGTGCGCTTGAGC 64507
QY 235 TGTTCAGAGCGCGGCTC---GACGTGTCTGAGGAGG---CCCGAGGGGGGGGAGGTG 290
Db 64508 TGTTCAGAGCGCGGCTCGATGACAGGTCTGAGGAGGCGCCAGAGCGCGGGAAGTG 64567
QY 291 GCCCAGACAGCGGGTCTGTAAGAGAGCGTTGGGAAGATTCGATCCGAGAGAGAGAA 350
Db 64568 GCCCGAGAGCGGGTCTGTAAGAGACGTTGAGAGAGATTCGATCCGAGAGACAGAA 64627
QY 351 GAACCGGATTGAAAGAGAGCGCGCTGAGGAGGAGGGGCTCTCTAAGATGGCGTCGG 410
Db 64628 GAGCGGATGAAAGAGACAGCGCGCTGCGGGAAGGGGCTGCCAAGATGGCGTCGG 64687
QY 411 CTCTCTCGGGCGCTGCTCTTCGGTGGTGTTCATCTTTGATCCCGCGGCTCCCTTCCT 470
Db 64688 CTCTCTCCCAACCGCTGTTGGCGGTGTTTTTTCATCTTTGATCCCGGGGCGCTTCCT 64747
QY 471 GTACCTTGCTCTCAGCATCTGGAATCAAGAGACCCATGGCATCTGAGTGCTTTATGCT 530
Db 64748 GTACCGGCTCTCAGCATCTGGAATCTTGAGCCCGACGGCATCTGAGTGCTTTATGCT 64807
QY 531 CTGSCATGCCCATCAAGAAATAGGCCATAGAACTGTTGATTCTCTCAGGAGAGACAACAT 590
Db 64808 CTGSCATGCCCATCAAGAAACAGGCCATCGAGGTGTCGATTCCTCTCAGGAGAGACAACAT 64867
QY 591 ATAAAAGACAACCTCATACGCTTGAAGGTGCCATCCAGTTAGGCATATACCCACTGTG 650
Db 64868 ATAAAAGACAACCTCAACAGCCTTGAAGGTGCCATCCAGTTAGGCATATACCTACTG 64927
QY 651 TGGGGACCTGAGTACCACCAACAGAGCGTGTCTCTCATGCAAGATTTCTACCTGGTGG 710
Db 64928 TGGGGACCTGAGTACCACCAACAGAGCGTGTCTCTCATGCAAGATTTCTACCTGGTGG 64987
QY 711 AGAGTATCTCTTTCCAGTGAAGGAGCAACCTGACCCCTGCCTCATCATACTACATGCT 770
Db 64988 AGAGTATCTCTTTCCAGTGAAGGAGCAACCTGACCCCTGCCTCATCATACTACATGCT 65047
```


Db 67196 TTATTACTGTTTGAATTTGAATATATCTGCTGCTTTCTAAATATCAAGACTTAC 67255
Qy 2967 CAATGAATTTAGATCATCTCCAGAGGAGATTTTTTTCCTCTCTCATCTTTTCCAA 3026
Db 67256 CAATGAATTTAGATCATCTCCAGAGGAGATTTTTTTCCTCTCTCATCTTTTCCAA 67314
Qy 3027 CAGTGTCTCTCTGTTGAGGCTTAAGTAAAGAGGGGACACTTCTCTGTTTAAACAGA 3086
Db 67315 CAGTGTCTCTCTGTTGAGGCTTAAGTAAAGAGGGGACACTTCTCTGTTTAAACAGG 67374
Qy 3087 CAGTCCATATCTGAGGCGCAGCAATATTTTCTTAAACATCATGGGAGAGCAGATTC 3146
Db 67375 CAGTCCATATCTGAGGCGCAGCAATATTTTCTTAAATTCATGGGAGACAGTGGATTC 67434
Qy 3147 TTGCTTGTGTGAGTCACTGCTGCGCATATGCTTACCCCTCTCTTATGCGAGGAA 3206
Db 67435 TTGCTTGTGTGAGTCACTGCTGCGCATATGCTTACCCCTCTCTTATGCGAGGAA 67492
Qy 3207 GTTGGAAATGGGGCTACATATGCCCCTCTCTCCCTCTACAGAGTGTGTTTCCCA 3266
Db 67493 GTTGGAAATGGGGCTACATATGCCCCTCTCTCCCTCTACAGAGTGTGTTTCCCA 67552
Qy 3267 TCTGATCCTTCCACTCTGTGAGGGAAGAGGGGGCTTGTATCTCAGGAGATTTGTTG 3326
Db 67553 TCTGATCCTTCCACTCTGTGAGGGAAGAGGGGGCTTGTATCTCAGGAGATTTGTTG 67612
Qy 3327 AATTCCTGTTCTATCCCTCTCTATCCACCCTGCTTGTATATATGTTAGCCCATACC 3386
Db 67613 AATTCCTGTTCTATCCCTCTCTATCCACCCTGCTTGTATATATGTTAGCCCATACC 67672
Qy 3387 CAATAACTGTCTATATAGACACCCCTGAGTCTGCTGCTCTCTTCTGCTGCTGCTG 3446
Db 67673 CAATAACTGTCTATATAGACACCCCTGAGTCTGCTGCTCTCTTCTGCTGCTGCTG 67732
Qy 3447 TG----TTTTTACAAGGAAGAAATCTGTGCT--ATTTTTTTTTCATATTTACTATT 3501
Db 67733 TGTTCTTTTACAAGGAAGAAATCTGTGCTATTTTCTTCTATATTTTCTATATTT 67792
Qy 3502 TATGATCTATTAGTCTTTTATTAAGACAGAGTCTCTGTTAGGGTGGGAGGAATATT 3561
Db 67793 TATGATCTATTAGTCTTTTATTAAGACAGAGTCTCTGTTAGGGTGGGAGGAATATT 67852
Qy 3562 TGAGGAGGGCTGCTGCTTGAAGGAAGAAATGGAAGCAACATTTTATTAAGTGTATC 3621
Db 67853 TGAGGAGGGCTGCTGCTTGAAGGAAGAAATGGAAGCAACATTTTATTAAGTGTATC 67912
Qy 3622 TATTGCTCTACTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3681
Db 67913 TATTGCTCTACTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 67972
Qy 3682 TTTTAAATGTAATAACTTTTAAATGTTATTTA 3713
Db 67973 TTTTAAATGTAATAACTTTTAAATGTTATTTA 68004

RESULT 4
BC028580
LOCUS BC028580 4262 bp mRNA linear PRI 25-APR-2002
DEFINITION Homo sapiens, similar to phosphatidylinositol-4-phosphate 5-kinase,
type I, alpha, clone MGC:26597 IMAGE:4828163, mRNA, complete cds.
ACCESSION BC028580
VERSION BC028580.1 GI:20306245
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 4262)
REFERENCE Strausberg R.
AUTHORS Direct Submission
TITLE Submitted (23-APR-2002) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK
COMMENT

USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-email.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 34 Row: d Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Genomescan gene prediction, Similarity but not identity
to protein.

FEATURES
source

Location/Qualifiers

1. 4262

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:26597 IMAGE:4828163"

/tissue_type="Testis"

/clone_lib="NIH_MGC_97"

/lab_host="DH10B"

/note="Vector: pBluescript"

440..1978

/codon_start=1

/product="Similar to phosphatidylinositol-4-phosphate

5-kinase, type I, alpha"

/protein_id="AAH28580.1"

/db_xref="GI:20306246"

/translation="MASASSPSLAVGFSFDPGAPCTASSAGSILPTASEVPYAS

GMPIKKTGHEGVDSSGETTYKTKTALKGAIQIGITTYVGSISTKPERDVLQMDFYV

VESIFPSEGSNLTPAHYNAPFRKTAPVAFYRFELFGIPDDYSLSCSEPLIEL

CSSGAGSLFYVSSDELIITLQHEAEFLQKLPGLYLNLSQNPRTLLPFFGLYC

YQRTGKNRIVVMNLLIPRSVKHIDKAGSYKRRASQKREKPLPTFDLDFLDQ

IPDLFLDADTYNALCTLQDCLVLQSFIMDYSLWLSIHNTDHAGREPLSSDTLQV

SIDQRLAPOKALYSTAMEFIQGEARLQDMEADHMGIPNASKGERLLLYIGIID

ILQSYTFLLKLEHSAKVAHVHDGDAVSVHRPFAERFQHFMCNFAFKKIPKPSPK

FRGLSFLSHLTSGNSCTTQPLVSEHKSQVQIKVQVEPGVHLGRSDVLPTQSESTF

GGNQ"

BASE COUNT 1117 a 981 c 990 g 1174 t

ORIGIN

Query Match 81.8%; Score 3037.6; DB 9; Length 4262;

Best Local Similarity 91.1%; Pred. No. 0;

Matches 3419; Conservative 0; Mismatches 279; Indels 54; Gaps 16;

Qy 2 TTAACAGGCGTGTGTAGGAAGGACGAGAGGGGCGCTCGCTCTTTGGGACCTTTTCAT 61

Db 30 TAAACAGGCGTGTGTAGGAAGGACGAGAGGGGCGCTCGCTCTTTGGGACCTTTTCAC 89

Qy 62 GCCTCG-----TTTTTTTTCAGATGTGGCTTTGGTCTGGGCGCAAGGTCCCAAGCACC 114

Db 90 GCCTCGTTTGTGTTTCTTCTCAGATGTGGCTTTTCTCAGCAGCGGTCCCAAGCAGTC 149

Qy 115 AGCTTAAGCTTACTTCTCTGTGAAAGGGGAAAGTATCCCTCTGTGAAAGCGGTTAAACTT 174

Db 150 AGCTTGAGCTTACTTCTCTGTGAAAGCGGAAAGTCTCCTCTGAGGAAAGCGGTTGAAC 209

Qy 175 GTGAGGCGGTGCGGAGCGGTGAGTCTTCCCATGCCAGGCGAATGGTGGCTTGGAGC 234

Db 210 GTGAGGCGGTGCGGAGCGGTGAGTCTTCCCATGCCAGGCGAATGGTGGCTTGGAGC 269

Qy 235 TGGTCCAGGAGCGGCTC---GACGTGTCTGAGGAGG-CCCGAGGGGGGGGAGGTG 290

Db 270 TGGTCCAGGAGCGGCTC---GACGTGTCTGAGGAGG-CCCGAGGGGGGGGAGGTG 329

QY 291 GCCCAGAGACCGGGTCTGTAAGAGACGTTGGGAAGATTCGATTCGCGAGAGAGAA 350
DB 330 GCCCGAGAACCGGGTCTGTAAGAGACGTTGGGAAGATTCGATTCGCGAGAGAGAA 389
QY 351 GAACCGGATTGAAGAGACCGAGCGCTGAGGGAGGGGGCTGCTAAGATGGCGTGG 410
DB 390 GAGCGGGATTGAAGAGACCGCGGGGGCTGCGGGGAAGGGGGCTGCCAAGATGGCGTGG 449
QY 411 CTCTCTCGGGCGGCTGCTGCTCGGTCTGTTTTCATCTCTTTCATCCCGGGTCCCTTCT 470
DB 450 CTCTCTCCCAACCGCTGCTGCGGGTGGTCTGTTTTCATCTCTTTCATCCCGGGGGCTCTCT 509
QY 471 GTACCTTGTCTCAGCATCTGGAATCAAGAGACCCATGSCATCTCAGGTGCTTATGCT 530
DB 510 GTACCGCATCTCAGCATCTGGAATCTTGAGCCCCACGGCATCTCAGGTGCTTATGCT 569
QY 531 CTGGCATGCCCATCAAGAAATAGCCCATAGAAGTGTGATTTCTTCAGGAGAGACAAT 590
DB 570 CTGGCATGCCCATCAAGAAACAGGCCATCTCAGGTGTGATTTCTTCAGGAGAGACAAT 629
QY 591 ATAAAAAGACAACCTCATCAGCCTTGAAGGTGCCATCCAGTTAGGCATTTACCCACACTG 650
DB 630 ATAAAAAGACAACCTCAACAGCCTTGAAGGTGCCATCCAGTTAGGCATTTACACTG 689
QY 651 TGGGAGCCTGAGTACCAAAACAGAGCGGTGATGTCTCATGCAAGATTTCTACGTGGTGG 710
DB 690 TGGGAGCCTGAGTACCAAAACAGAGCGGTGATGTCTCATGCAAGATTTCTACGTGGTGG 749
QY 711 AGAGTATCTCTTCCACGTGAAGGAGCAACCTGACCCCTGCTCATCACTACAATGACT 770
DB 750 AGAGTATCTCTTCCACGTGAAGGAGCAACCTGACCCCTGCTCATCACTACAATGACT 809
QY 771 TTCGTTTCAAGACCTATGACCTGTTGCTTCCGCTTACCTTCGGGAGCTATTTGGTATCC 830
DB 810 TCCGTTTCAAGACCTATGCGCGGTGCTTCCGCTTACCTTCGGGAGCTATTTGGTATCC 869
QY 831 GGCCCGATGATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 890
DB 870 CGCCCGATGATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 929
QY 891 GAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 950
DB 930 GAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 989
QY 951 AACATAAAGAGCGGGAATTTCTGCAAGAGCTGCTTCCAGGATCTACATGAACCTCAACC 1010
DB 990 AACATAAAGAGCGGGAATTTCTGCAAGAGCTGCTTCCAGGATCTACATGAACCTCAACC 1049
QY 1011 AGAACCTCGGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1070
DB 1050 AGAACCTCGGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1109
QY 1071 AGAACATTCGGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1130
DB 1110 AGAACATTCGGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1169
QY 1131 AATATGACCTCAAGGCTCAACCTACAAACGGCGGCTTCCAGAAAGAGCGAGAGAAC 1190
DB 1170 AATATGACCTCAAGGCTCAACCTACAAACGGCGGCTTCCAGAAAGAGCGAGAGAAC 1229
QY 1191 CTCTTCCACATTAAGACCTAGACTTCTTACAGACATCCCTGATGCTTTTTTTGG 1250
DB 1230 CTCTTCCACATTAAGACCTAGACTTCTTACAGACATCCCTGATGCTTTTTTTGG 1289
QY 1251 ATGCTGACATGTACAACGCTCTGTAAGACCTGCTGAGCGGTGACTGTTTGGTCTGCTGACA 1310
DB 1290 ATGCTGACATGTACAACGCTCTGTAAGACCTGCTGAGCGGTGACTGTTTGGTCTGCTGACA 1349
QY 1311 GCTTCAAGATAATGGATTACAGCCTTCTGATGCTCAATTCATATATGATCATGCAACAAC 1370
DB 1350 GCTTCAAGATAATGGATTACAGCCTTCTGATGCTCAATTCATATATGATCATGCAACAAC 1409
QY 1371 GAGAGCCCTTAAGCATGTAACACAGATA ---CTCAGTTGATCTCGAAGACCGGCCCCC 1427

DB 1410 GAGAGCCCTTAAGCAGCGACACTCTTCAAGTGTCAATCGACACTCAAAAGACTGGCTCCCC 1469
QY 1428 AAAAGGCTCTGATTTCCACAGCCATGGAAATCCATCCAGGGAGAGGCTCGAGCGGTGGTA 1487
DB 1470 AAAAGGCTCTGATTTCCACAGCCATGGAAATTCATCCAGGGAGAGGCTCGGTGGCGGACA 1529
QY 1488 CCATGGAGACTGATGACCATATGGGTGGCATCTCCCTGCCCGGGAATAGTAAGAGGGAAGGCG 1547
DB 1530 CCATGGAGCGCATGACCATATGGGTGGCATCTCCCTGCTCAGATAGTAAGAGGGAAGGCG 1589
QY 1548 TTCTGCTTTATTTGGCATCATTTGACATTTACAGTCTTACAGGTTTCTTAAAGATTGG 1607
DB 1590 TTCTGCTTTATTTGGCATCATTTGACATTTACAGTCTTACAGGTTTCTTAAAGATTGG 1649
QY 1608 AGCATCTTTGAAAGCGCTGTTACATGACGAGAGACACGTCTCAGTGTGATCGCCAGGCT 1667
DB 1650 AGCATCTTTGAAAGCGCTGTTACATGATGGGAGCGTGTCTCAGTGTGATCGCCAGGCT 1709
QY 1668 TCTACGCTGAACGGTTCCAGCGCTTCTATGCAACACAGTATTTAAGAGATTTCCCTTGA 1727
DB 1710 TCTACGCTGAACGGTTCCAGCGCTTCTATGCAACGCACTATTTAAGAGATTTCCCTTGA 1769
QY 1728 AGCTTCTCTTCCAAAAGTTTGGTCTGGCTCATCTTCTCTCGGCGAGGAGGCTCCA 1787
DB 1770 AGCTTCTCTTCCAAAAGTTTGGTCTGGCTCATCTTCTCTCTGATACGGGCTCCA 1829
QY 1788 GTGGCAACTCTGTCATTTACTTACCAGCCATCGTCTCTGGGGAACACAAAGGCACAAAGTGA 1847
DB 1830 GTGGCAACTCTGTCATTTACTTACCAGCCATCGTCTCTGGGGAACACAAAGTGA - 1888
QY 1848 CAACAAAGGCAAGAGTGGAGCGCTTACCTTTGCTGCTGCTGATGTTTACCTCAGA 1907
DB 1889 --ATAAGGCTGCAAGTGGAGCGAGTGTACCTTTGGTCTGCTGATGTTTACCTCAGA 1946
QY 1908 C-----TCCACCTTTGGAGGAATAGTGGAGGCTCGCTTATTTCCCTACCCAGTTTC 1960
DB 1947 COTCAGATCCACCTTTGGAGGAATCAGTGGAGGCTC-ACTATTACTGACCCACAGTTTC 2005
QY 1961 TCACCTCTAGTTGGAGAGACTTTGCAAAATGCTAACTACAAGTACAACTTTGGAAGAGTT 2020
DB 2006 TCACCTCTAGTTGGAGAGACTTTGCAATGCTAACTACAAGTATTAACCTTTGGAAGAGTT 2065
QY 2021 GAAGTTGAGAGTCAAGTTTACCATTAAAGCGCAAGGCTCAGAAAGACCTTGGAAACAAGA 2080
DB 2066 GAATGTACAGAGTCAAGTTTACCATTAAAGCGCAAGGCTCAGAAAGACCTTGGAAACAAGA 2125
QY 2081 TTCTGCGCATCTGATGCCCAAGATGTCAGCCCTTGGCCCAAGCAATGCTGAATTTCTT 2140
DB 2126 TTCTG-CTTCTGTTGATCCCAAGATGTCAGCCCTTGGCCCAAGCAATGCTGAATTTCTT 2184
QY 2141 CTACTTGGTCAATCAAAAAGGAGTGAATAGAAGTGGAGGAGCTGCTCTCATCTTCT 2200
DB 2185 CTACTTGGTCAATCAAAAAGGAGTGAATAGAAGGAGGAGTGCACCT---CTCATCTTCT 2240
QY 2201 TCTTGAAGAGAACCTTCT 2260
DB 2241 TCTTGAAGAGAACCTTCT 2300
QY 2261 AGTTGAGGA- -----CCGAGCATCCCTCCACTCCAGAGTGGGTGCTAGG 2306
DB 2301 AGTTGAGGAGAACCTTCT 2360
QY 2307 GATTTTCAACTGCGCAACCTTTGCTCTCACTTAATTTTCACTTTTTCAGACCCCTTCTT 2366
DB 2361 GATTTTCAACTGCGTAA-CCTTGGCTTCCACAACCTGAATTTTTCAGACCCCTTCTT 2419
QY 2367 CATGCTGGAATGGAGTTGCTGGAATTTGGAGCTTCTTCTTCCCTGCTCTTTGACTAGA 2426
DB 2420 CATGCTGGAATGGAGTTGCTGGAATTTGGAGCTTCTTCTTCCCTCATCTTTTTCAGCAGA 2479
QY 2427 ACCGAGCTCTTAATTTCTCTCAGGAGAGACTAGCTGCGACATTTACCTTACCTTCTTCT 2486

DB 158002 TTTTATGTAATAAACTTTTATCAGTTATTTA 158033
 |||||
 AC006476 112804 bp DNA linear PRI 30-SEP-2000
 LOCUS Homo sapiens PAC clone RP4-764012 from 7, complete sequence.
 DEFINITION AC006476
 ACCESSION AC006476
 VERSION AC006476.3 GI:4926910
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 112804)
 Sulston, J.E. and Waterston, R.
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 99063792
 9847074
 2 (bases 1 to 112804)
 Cordes, M., Wohlman, P. and Phillips, A.
 The sequence of Homo sapiens PAC clone RP4-764012
 Unpublished
 3 (bases 1 to 112804)
 Waterston, R.H.
 Direct Submission
 Submitted (30-JAN-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 112804)
 Waterston, R.H.
 Direct Submission
 Submitted (29-MAY-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 112804)
 Waterston, R.
 Direct Submission
 Submitted (30-SEP-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On May 29, 1999 this sequence version replaced gi:4337287.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 ----- Center project name: H_DU0764012

 NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.
 This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.
 MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and
 sequencing collaboration between the NHGRI Chromosome 7 Mapping
 Project (Eric D. Green, Director), John D. McPherson in the
 Department of Genetics (Washington University), and the Washington
 University Genome Sequencing Center. For additional information
 about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send

mailto:egreen@nhgri.nih.gov , or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 This clone was derived from human PAC library RPCI-4, prepared by
 Pieter de Jong and coworkers at the Roswell Park Cancer Institute
 (<http://pacpac.med.buffalo.edu>) using the method described by
 Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from
 one male donor.
 The clone may be obtained either from Genome Systems, Inc.
 (<http://www.genomesystems.com>) or Research Genetics, Inc.
 (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2
 NEIGHBORING SEQUENCE INFORMATION:
 A tandem repeat, extending from base position 67681 to 71791, has
 been sorted to best possible assembly. Restriction digest
 information with bamhi, hindiii, and ecorv suggests this region is
 missing approximately 1400 base pairs of sequence.

The clone sequenced to the left is RP5-1058P19. Actual start of
 this clone is at base position 1 of RP4-764012; actual end is at
 112804 of RP4-764012.

FEATURES	Location/Qualifiers
Source	1..112804
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7"
	/clone="RP4-764012"
	/clone_lib="RPCI-4"
repeat_region	1..65
	/rpt_family="Alu"
repeat_region	282..668
	/rpt_family="MaLR"
repeat_region	1575..1615
	/rpt_family="AT-rich"
repeat_region	1616..1897
	/rpt_family="Alu"
repeat_region	3392..3775
	/rpt_family="Retroviral"
misc_feature	4319..4868
	/note="match to EST AI568539 (NID:94531913) tn40e08.xl"
repeat_region	5673..5921
	/rpt_family="MER1_type"
misc_feature	8043..10028
	/note="CpG island (%GC=68.4, o/e=0.90, #CpGs=214)"
repeat_region	8332..8364
	/rpt_family="A-rich"
repeat_region	9202..9330
	/rpt_family="(CCCCG)n"
repeat_region	9950..9970
	/rpt_family="GC-rich"
repeat_region	10090..10212
	/rpt_family="L1"
repeat_region	10986..11122
	/rpt_family="(TAGA)n"
repeat_region	11243..11360
	/rpt_family="(CAGA)n"
repeat_region	12015..12240
	/rpt_family="MER4-group"
repeat_region	12243..12853
	/rpt_family="MER4-group"
repeat_region	12881..12991
	/rpt_family="MER94"
repeat_region	12998..13524
	/rpt_family="L1"
repeat_region	16196..16425
	/rpt_family="Retroviral"
repeat_region	17256..17613
	/rpt_family="Retroviral"
repeat_region	18007..18149
	/rpt_family="MER2_type"
repeat_region	18224..18522
	/rpt_family="MER2_type"

```
repeat_region 18715..18832
/rpt_family="L1"
repeat_region 19286..19373
/rpt_family="L2"
repeat_region 20391..20481
/rpt_family="MIR"
repeat_region 21808..22111
/rpt_family="Alu"
repeat_region 22147..22188
/rpt_family="At-rich"
repeat_region 22392..22555
/rpt_family="MaLR"
repeat_region 23127..23212
/rpt_family="MaLR"
repeat_region 24894..25083
/rpt_family="MER4-group"
repeat_region 25147..25237
/rpt_family="MER4-group"
repeat_region 25269..25598
/rpt_family="MER4-group"
repeat_region 25816..26607
/rpt_family="MER4-group"
repeat_region 26945..27063
/rpt_family="MER4-group"
repeat_region 27210..27502
/rpt_family="Alu"
repeat_region 27503..27768
/rpt_family="MER4-group"
repeat_region 28822..29089
/rpt_family="MER4-group"
repeat_region 29111..29368
/rpt_family="MER4-group?"
repeat_region 29377..29464
/rpt_family="(CATAT)n"
repeat_region 29748..29786
/rpt_family="At-rich"
repeat_region 29893..30466
/rpt_family="L1"
repeat_region 30524..30638
/rpt_family="(CATA)n"
repeat_region 30886..30997
/rpt_family="(CA)n"
repeat_region 31015..31125
/rpt_family="(CAATA)n"
repeat_region 31099..31230
/rpt_family="(CACAC)n"
repeat_region 32285..33146
/rpt_family="L1"
repeat_region 33240..33373
/rpt_family="MER2_type"
repeat_region 33380..33559
/rpt_family="L1"
misc_feature 35690..37077
/Note="CpG island (%C=63.0, o/e=0.90, #CpGs=106)"
repeat_region 37890..40107
/rpt_family="MER2_type"
repeat_region 42804..43109
/rpt_family="Alu"
repeat_region 44304..44361
/rpt_family="MIR"

Query Match 56.18; Score 2081.2; DB 9; Length 112804;
Best Local Similarity 75.3; Pred. No. 0;
Matches 3248; Conservative 0; Mismatches 403; Indels 665; Gaps 26;

Qy 20 GAAGACGAGAGGGGGTTCGCTTTGGGACTTTTCAGCTTCCTTTTTTTTTCAG 79
Db 98250 GGAGACGAGAGGGGGGCTTCACTCCGCTGGGGTTCATGCGACGCTTTTCTCCAG 98309

Qy 80 ATGTGGCTTGTCTGGCGCAAGGTCCACAGCAGCAGCTTAAGCTTACTCTCTGTGAAA 139
Db 98310 GTGTGACTTGTCT--GGACGCGGATCCGGCAGCAGCTGGGGCTTAGTCTCTGTGAAA 98367
```

```
Qy 140 GGGAAAGATATCCCTGTGGAAAGGGGTTAAACTTGTGGAGGGGTGGGGACGTGAGTT 199
Db 98368 GAGTAAAGTGCCTCCTCGGAGAGGGGTAACCTGCCAAGGGCTGCGAGCGGAGGTT 98427

Qy 200 CTTCCCATGCAGGCGGAATGTTGGCTTTGAGCTGTGTCAGGAGCGGGCTCAGACGTGT 259
Db 98428 C-TCCCCAGGCGAGCGAATGGGCGGCGCTGAGCTGTCTA-GAGCGGGCTCCACGGCT 98485

Qy 260 CTGAGGAGGCGGAGGGGCGGGAGGTGGCCACAGAACCGGATGAAAGAGACGACGCC 319
Db 98486 CTGAGGCGGCGCGGGGCGGGAGGTGGCCCA---GACGGGGTCTCTGTAAGGGA 98542

Qy 320 GTTTGGGAAG--ATTGATTCGAGAGAGAGAAAGCGGATGAAAGAG-GCCAGGCC 376
Db 98543 CTTGGGGAAGGAAGATTCATTCGAGAGAGAAAGCGGATGAAAGAG-GCCAGGCC 98601

Qy 377 GCTGAGGAGGAGGGGCTGCTAAGATGGCGTGGCGCTCCTCCGGGCGCTCGTCTCGGTC 436
Db 98602 GCTGAGGGG--AGGGGCTGCCAAGATGGCGCTCGTCTCCTCAGGCGCTCCTCGGCGTC 98659

Qy 437 GGTTCATCCTTTGATCCCGCGGTCCCTTCTGTGTAAGTGTGCTCAGCATCTGGAATC 496
Db 98660 AGTTTTCATCCTTTGATCCAGGGGTCGCTTCTGTACCTGCTCAGCATCTGGAATC 98719

Qy 497 AAGAGCCCATGGCATCTGAGTGCCTTATGCCTGTGCATGCCCATCAAGAAATAGGC 556
Db 98720 CAGAGCCTATGGCGTCTGAGTGCCTTGTGCCCTGGGCATACCCATCAAGAAATAGGC 98779

Qy 557 CATAGAAGTGTGATTCCTCAGGAGAGACAACATATAAAAGACAACCTCATCAGCCTTG 616
Db 98780 CATCGAGTGTGATTCCTCAGGAGAGACAACCTATAGAAAGACGACCTCATCAGCCTTG 98839

Qy 617 AAAGTGCCATCCAGTTAGGCATTACCCACACTGTGGGAGCGCTGAGTACCACACAGAG 676
Db 98840 AAATGTGCATCCAGTTGTGCATTAACACTGTGGGAGCGCTGCATCCACACACAGAG 98899

Qy 677 CGTATGTCCTCATGCAAGATTTCTACGTGTGAGAGTACTTCTTTCCAGTGAAGG 736
Db 98900 CGTATGTCCTCATGCAAGATTTCTCGTGTGGGAGAAATATCTTTCCAGTGAAGG 98959

Qy 737 AGCAACCTGACCCCTGCTCATCACTACAATGACTTTCCTTTCAAGACCTATGCACTGT 796
Db 98960 AGCAACCTGACCCCTGCTCATCACTGCACTGAGTTCTTTCAAGACCTATACACTGT 99019

Qy 797 GCCTTCGCTACTTCGCGGAGCTATTGCTGATCCGGCCGATGATGATTTTCCTC 856
Db 99020 GCCTTCGCTACTTCGAGGAGCTATTGCTGATCTGCCCTGATGATGATGATCTCTCCTC 99079

Qy 857 TGCAGTGAGCGCTGATTGAACCTCTGTAGCTCTGGAGCTAGTGTCCCTATCTATCTG 916
Db 99080 TGCAGTGAGCGCTGATTGAACCTCAGTAGCTCTGGAGCTGATGGTTCCTGTTACATGTG 99139

Qy 917 TCCAGCGAGATGAGTTTCATTATTAAAGACAGTCCAACATAAAGAGCGGAAATTTCTGCG 976
Db 99140 TCCATCCACATGAATTCATTATTAAACAGTCCCAACATAAACAGCGGAGTTCTGCGAG 99199

Qy 977 AAGCTGCTTCAGGATACATGAACCTCAACAGAACCTCGGACCTTGTGCTCTAAA 1036
Db 99200 AAGCTGATTCAGGATACACATCGACCTCAACAGAACTCTTGGACTTGTGCTCTAAA 99259

Qy 1037 TTCTATGAGCTGACTGTGTCAGGAGGTGGCAGAGACATTCGGATTTGTGGTGATGAC 1096
Db 99260 TTCTATGAGCTGCTGTGTGAAGCAGGTGGCAGAGACACGAGATTTGCGGTGATGAG 99319

Qy 1097 AATCTTTTACCAGATCGTAAAATGCATATCAAAATGACCTCAAAAGGCTCAACCTAC 1156
Db 99320 AATCTTTTACTAAGATGCTCAAAATGCATACCAATATGACTTCAAAAGTTCAACCTAC 99379

Qy 1157 AAACGGCGGGTTCAGGAGAGAGAGAGAGAGGCTC-----TCCACATTTAA 1206
Db 99380 AAAGAGCAGGCTTCAGGAGAGAGAGAGAGAGGCTCTCCACATTTTCCACATTTAA 99439

Qy 1207 AGACCTAGACTCTTACAGACATCCCTGTGTTGGTCTTTTTTGGATGCTGACATGTACAA 1266
```


Db	99440	AAACCTAGACTTCTTAAAGACATCCCGATGGTCTTTTTTTTGGATGCTAAAGATGTACAA	99499
Qy	1267	CGCTCTCTGTAAGACCCCTCAGCGGTGACTGTGTTGGTGCTGCAGAGCTTCCAAGATAATGGA	1326
Db	99500	CGCTCTCTGTAAGACCCCTCAGCGGTGCTGTTGGTGCTGCAGATCTTTGAGATAATGGA	99559
Qy	1327	TTACAGGCTCTTGATGTCAATCCATTAATATAGATCATGCACACAGAGAGCCCTTAAGCAG	1386
Db	99560	CTATAGGCTCTTGATGTCAATCCACAACATAGATCATGC---ACGAGAGCCCTTTGAGCAG	99616
Qy	1387	TGAACACAGACTACTCAGTTCTGATCTCGAAGACCGGGCCCCCAAAAGGCTCTGTATTCCAC	1446
Db	99617	CCAAACAAATGCTCAGTTACACTTTGAAGACACGCCCCCAAGGGCTCTGTATTCCAC	99676
Qy	1447	AGCCATGGAAATCCATCCAGGAGAGGCTCGACGGGTGGTACCATCGAGACTGATGACCA	1506
Db	99677	AGCCATGGAAATCCATCCAGGAGAGGCTTGCAGGGGAGCCCATCGAGACTGATGACCA	99736
Qy	1507	TATGGGTGGCATCCCTGCCCGGAATAGTAAAGGGAAAGGCTTCTGCTTTATATTGGCAT	1566
Db	99737	TATGGGTGGCATCCCTGCCCAGAAATAGTAGGGGGAAGAAATCCTCTGTTGATTGGCAT	99796
Qy	1567	CATTGACATTTCTACAGTTCTTACAGGTTTGTTAAGAAGTTGGAGCACTCTTGGAAAGCCCT	1626
Db	99797	CACGTGACATTTCTACAGTTCTTACAGGTTTGTTAAGAAGCTGGAGCACTCTTGGAAAGCCCT	99856
Qy	1627	GGTACATGACGGAGACACTGTCTCAGTGCATCCCGAGGCTTCTACGCTGAACGGTTCCA	1686
Db	99857	GGTATGTATGGGAGCACTGTCTCAGTGCATCCCGAGGCTTCTACGCTGAACGGTTCCA	99916
Qy	1687	GCGTTCATGTGCACACAGTATTTAAGAAAGATT-----	1720
Db	99917	GTGCTTCATGTGCAACACAGCATTTTAAAGAAGATCCCGGCCAGGTGCGGTGCTCGCACT	99976
Qy	1721	-----	1720
Db	99977	GTAATCCAGACTTTTGGGAGGCCAAGCGGATGGATCACTCAGGTCAAGGTTCAAGA	100036
Qy	1721	-----	1720
Db	100037	CCAGCTTGGCAACATGGTGAACCCCATCTCTACTAATAATCACAAATATGCGCGGGG	100096
Qy	1721	-----	1720
Db	100097	GTAGTGGACATGCTGTGAACCTCAGCTACTTGTGTAGGCCAAGCAGGAGAAATCACTTGA	100156
Qy	1721	-----	1720
Db	100157	GCCTAGGAGCAGAGTTGCAGTGAGCCAAGATTGCAACCACTGCATCCAGCCTTGGGCAA	100216
Qy	1721	-----CCCTTTGAAGCCTTCTCTCTTCCCAA	1744
Db	100217	AAGAGTGAGCCCTGTCTCAAAAAGAGAAGATCCCTTTGAAGCCTTCTCTCTTCCAA	100276
Qy	1745	AAGTTTCGGTCTGCTCATCTTTCTCTCGGAGCAGAGGTCCAGTGGCAACTCCTGCATT	1804
Db	100277	AAGTTTCGGTCTGCTCATCTTTCTCTGCGGAGCAGGCTCCAGTCAACTCTCTG	100332
Qy	1805	ACTTACAGCCATCGGTCTCTGGGGAACACAAGGCAACAAGTACAAACAAGGCAAGATG	1864
Db	100333	TGTTACAGCCATCGGTCTCAGGGGAACACAAGGCAACAACCAACAACAAGGCGGAAGTG	100392
Qy	1865	GAGCCAGCGTTACCTTGGTCTGCTGATGTTTACCCTCAGACTCCACCTTTGGAGGAA	1924
Db	100393	GAGCCAGGTGTACCTTGGTGTGCTGATGTTTACCCTCAGACTCCACCTTTGGAGGAA	100452
Qy	1925	ATCAGTAGGCGCTCGCCTATTCTCTGAGCCAGTTCTCAGCTCTAGTTGGAGAGACTTGTG	1984
Db	100453	ATCAGTAGGCGTTTACCTATTCTCTGAGCCAGTTCTCAGCTCTAGTTGGAGAGACTTGTG	100512
Qy	1985	CAATGCTAATACAGTACACCTTGGGAAAGCTTGAAGTTGCAGAGTCAGATTCACC	2044

Db	100513	CAAAATGCTAACCCACAAGTACACACCTTGGAAAGCGTTGACATTCGACAGAGTAAGAGTTACACC	100572
Qy	2045	CATTAAAGCGCAAAAGCCCTCAGAAAGACCTCGAAACAAGATTCTGCACATCTCTGTGATCCCAAG	2104
Db	100573	CATTGAGCACAAAGCCCTCAGAGACCTGAACAAGATTTCGCCATCTCTGTGATCCCAAG	100632
Qy	2105	ATGTCAGCCCTTGCCCCAGCAATGCTGAATTTTCTTCTTACTTTGGTTCATCAAAAAAGGAGT	2164
Db	100633	ATGTCAGCCCTTGCCCCAGCAATGCTGAGATTTTCTTCTTACTTTTCCCAT--TAAAAAGGAAT	100691
Qy	2165	GTAATAGAAGTGAGGGGAGCTGCTCTCCATCTTCTTCTCTGGAAGAACAACCTTCTCTCTCT	2224
Db	100692	GTAATGAAGGAGGGGAGCT---CTCCATCTTCTTCTCTGGAAGAAGAACCCTAGTCTCTCT	100747
Qy	2225	TCCTCTTCTCTCATGAATGGCGCTTAGTGCCCTCAGAGAGTTGAGGACCGCAGCATCCCCCTC	2284
Db	100748	TCCCTTCTCTCATGAATGGCGCTTAGTGCCCTCAGACAGTTGAGGACTCGAGCAACCCCTC	100807
Qy	2285	CACCTCCAGAGTTGGTGGTACGGATTTTCAACTGGCCCAACCCCTTTCGCCCTCCACTATTTGAA	2344
Db	100808	CACCTCCAGAAATTTGGTGG-ATGGATTTTCAATGGGCTAACCC--TTGGTCTTTCACAAATTGA	100864
Qy	2345	TTTTTTTTCAGACCCCCCATTTCTTCATGCTGGAATGGATTCGTGGACTTTGGCAGCTTTCT	2404
Db	100865	GCTTTTTCAGACCCCCCATTTCTTCATGCTGGGAATGGATTCGTGGACTTTGGCAGTTTCT	100924
Qy	2405	TTCCCTCTGCTTTGACTAGCAACCGGACHTTAATTTTCTCAGGACACAGCTAGCTGGCA	2464
Db	100925	TTCCCTCTGCTTTTCCAGGAGCTGGACHTCATTAATTTCTCAGGACACAGCTAGCTGGCA	100984
Qy	2465	CATTATCCCTACCTTAGTTTCTTCTCTGACRCCCTGGAGAATACTCTGTAATCTCTG	2524
Db	100985	CATTATACCCACCTCAGTTCTTCTTCTCTCATCCCTGGAAGAAGATCCCTGTAATCTCTG	101044
Qy	2525	TAAAGGTTTTTGGGGGATAGGGTGTTTTAACCACTCCCAAGCTTTCTTCTTCTTTTTTTT	2584
Db	101045	TAAAGGTTTTGAGGGGATAGGGTGTTTTAACCACTCCCAAGCTTTCTTCTTCTTTTTATCT-	101103
Qy	2585	TTCTGAAAAAGGAAAAAGCACACAGCACACAATTTCAAGCCATTTTCAGATCAGAACCTC	2644
Db	101104	--GAAATAAGGAAAAATGGCACAGCACGCCAATTTCAAGCCAGTTTCAGATCAAAACC	101160
Qy	2645	CAGAAGTTGTGACAAGATGCCTATTCTGTAGAGTTCCCTCAGAAGACCCATGGTGTATTG	2704
Db	101161	CAAAAGTTGTGACGAGATGCCTAGTCAATAGATTTCCTCAGAAGAACCATGGTGTATTG	101220
Qy	2705	AAGAGAAGTAGTAGTTGCTCTGCCAAGACAGCTCCCTTTTAAACTCCTCCTCTCTCTG	2764
Db	101221	AAGAGAAGTTGGTGTATTTGCTCTGCCAGGA-CAGCTCCTCTTTTAAACTCCTCCTCTCTG	101279
Qy	2765	ATGAATTTCTTAAAGCTGAAGGAATGAAGAGTGGGACATGGGTAACTTTATCTATCCCTT	2824
Db	101280	ATGAATTTGTAAAGCTTGGAGGAATGGAGAGTGGGACATGGGTAACTTTATCCCTT	101339
Qy	2825	TTGTTAAACAGGAGGACGCATGGGCTGGGAGATCATAGCCCTTCTTAGGCAGAACTCT	2884
Db	101340	TTTTTAAACATAGGGCAGCGGTGGCTGGGAGATCATAGCCCTTCTTAGGCAGAACTCT	101399
Qy	2885	GTTCACTGCCAGGCTATAGTAATTTACTATTGTCATTTTGAATATATATCTGGTGTG	2944
Db	101400	ACTCACTGCCAGGCTGTAATGATTTACTGTGTTTGGCAATTTGCAATATATATTC-GGTTGT	101458
Qy	2945	TTTTTCTAACTGGAAGACTTACCAAAATGATTTAGATCATTTCTCAGAGGAGATTTTTT	3004
Db	101459	TTTTTGTAATATGAAGACCTTATCAAAATTTTAGATCATTTCTCAGAGGAGATTTCTT	101518
Qy	3005	TTGCTCTTCTCATCTTTTCCAACAGTGTCTCTGTTGTGGAGCTTAAGGTAAGAGGGG	3064
Db	101519	TT-CTCTTCCATCTTTTCCAACAGTCTCTCTGCTGTGTGGAGCTAAGGTA-----GAG	101572
Qy	3065	ACACTCTGTCTGTTTAAACAGACAGTCCCATATCTGTGAGGCCAGCAA-----3111	
Db	101573	ACACTCTGTCTGTTTAAACAGGAGTCCATATCTATGAGGCTAGCAAAATTTTCTTTTC	101632


```
Qy 3112 ----- 3111
Db 101633 TTTCTTTTGAACAGTCTCACACTGTCGCTGGCTAGAGTGCAGTGGCGGATCTTG 101692
Qy 3112 ----- 3111
Db 101693 GCTCACTGCAACCTCCACCTCTCTGGTTCAAGCGATTCTCTGCTCAGCTCCCAAGTA 101752
Qy 3112 ----- 3111
Db 101753 GCTGGATTACAGTGCCTGCCACCATGCCAGCTAATTTTGTGTTTATAGTAGAGAT 101812
Qy 3112 ----- 3111
Db 101813 GGGGTTTCACTATGTTAGCAGGCTGGTCTCACTCTCGTGTATCCGCCACCTC 101872
Qy 3112 ----- 3119
Db 101873 GGGCGCCCAAGTCTGGGTTACAGGCGTGAGCCACCGCGCCGCCACAGATATTTC 101932
Qy 3120 TTAACCTCATCGGAGACAGAGATCTTGTGCTGGTGAGTCAATGCTGTGCCATATGT 3179
Db 101933 TTAACCTCATCGGAGACAGAGATCTTGTGCTCAGTCAATGCTGTGCCATATGT 101992
Qy 3180 CCTACCCCTCTCTTCATGACAGGAAGTTGGAATGGGGCTACATATGCCCTCTCCTC 3239
Db 101993 CCTA-CGCGCTGTCTTCATGACAGGAGGTTGGAATGGGGCGCCGATGTGCCCTCTCCTC- 102050
Qy 3240 CCCGCTCTACAAGAGTTGTGTTTCCATCTCATCTCTCCACTCTGTGCAGGGGAAGAAGG 3299
Db 102051 CCCATCTACAGAGTTGTGTTTCCATCTCATCTCTCCACTCTGTGTAGGGGAAGG 102110
Qy 3300 GGGCTGTATCTCAGGAGATGTTGAATTCCTGTTCTATCCCTCTCTATCCACCTC 3359
Db 102111 GGGCTGTATCTCAGGAGATGTTGAATTCCTGTTCTATCCCTCTCTATCCCTCTC 102166
Qy 3360 GCCGTGATAATATGTAGCCATACCCCAATACTGCTATATAGACACCCCGACCA 3419
Db 102167 ACCTTGATAGTAGGTAGCCATACCCCAATACTGCTATATAGACACCCCGACCA 102226
Qy 3420 GTTCTGCTGCTGCTCTCT-TTGCTGCCATGTTTTCACGAAGGAAG-AATCTTGCT 3477
Db 102227 GTTCTGCTGCTGCTCTCTGTTGCCATGTTCTTTTACGAAGGAAGAACATATT 102286
Qy 3478 ATTTTCTTTCATATTTACTATTTATGATGATTTAATAGTCTTTTATTAAGCACAGACT 3537
Db 102287 GCTATTTTTCATATTTACTATTTATGATGATTTAATAGTCTTTTATTAAGTATG 102346
Qy 3538 CTGTTAGGGTGGGAGGAATATTGAGGAGGGCTGGGTCTTAGGGAAGGAATGGGA 3597
Db 102347 CTGTAATGGTGGGAGGAATATTGAGGAGGGCTGGGTCTTAGGGAAGGAATGGGA 102406
Qy 3598 AGCAACATTTTATTAAGTGTACTATTTGCTCTCTACTTTCTGATTTGTCAGAAATGGCA 3657
Db 102407 ATCAACATTTATGAAGTGCTCTATTTGCTCTCTACTTTCTGATTTGTCAGAAATGGCA 102466
Qy 3658 ATACAATAAAGTGATATATGCTTTTAAATGTAATAACATTTATGATGATTTA 3713
Db 102467 ATGCAATAAAGTGTTACATGTTTAAATGATTTAAACATTTAATCATGATTTT 102522
```

```
RESULT 7
AF027390 AF027390 225432 bp DNA linear PRI 02-APR-1998
LOCUS Homo sapiens 7q telomere, complete sequence.
DEFINITION AF027390
ACCESSION AF027390
VERSION AF027390.1 GI:3004858
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
AUTHORS 1 (bases 1 to 225432)
Chi, H.-C., Saunders, E.H., Buckingham, J.M., Rieke, D.O., Munk, C.C.,
Lobb, R., Ueng, S.Y.-J., Mundt, M.O., White, P.S., Tatam, O.L.,
Riethman, H.C. and Moyzis, R.K.
TITLE DNA Sequence Analysis of the Terminal 226 kb of Human Chromosome 7q
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 225432)
Chi, H.-C., Saunders, E.H., Buckingham, J.M., Rieke, D.O., Munk, C.C.,
Lobb, R., Ueng, S.Y.-J., Mundt, M.O., Tatam, O.L., Riethman, H.C. and
Moyzis, R.K.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-1997) Center for Human Genome Studies, Los Alamos
National Laboratory, MS M888, New Mexico, NM 87545, USA
FEATURES
source
1. .225432
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q telomere"
901..1496
/note="HSAL000641"
/rpt_family="Alu"
complement(2341..2376)
/note="GRAIL 2 good exon ; ORF; frame 0"
2513..2800
/note="M32220"
/rpt_family="retrovirus"
2659..2997
/note="LTR2"
4101..4283
/note="HSAL006653"
/rpt_family="Alu"
complement(4512..4600)
/note="MER3"
/rpt_family="MER3"
8171..8421
/note="GRAIL 2 good exon ; ORF; frame 0"
8929..9106
/note="GRAIL 2 marginal exon ; ORF; frame 1"
12304..12466
/note="Z69721"
/rpt_family="L1"
complement(12460..12687)
/note="Z69364"
/rpt_family="Alu"
complement(12621..12766)
/note="97% identity -- U66083"
complement(12792..14659)
/rpt_family="L1"
complement(14751..14845)
/note="GRAIL 2 excellent exon ; ORF; frame 2"
19550..19636
/note="80% identity"
/db_xref="dbEST:R38951"
19655..19824
/note="80% identity -- Z77249"
19840..19998
/note="83% identity"
/db_xref="dbEST:W79023"
20128..20260
/note="81% identity -- Z77249"
21027..21097
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(24077..24350)
/note="U47924"
/rpt_family="Alu"
complement(24301..24492)
/note="HSAL000410"
/rpt_family="Alu"
24633..24901
/rpt_family="Alu"
complement(24992..25278)
/note="Z73359"
repeat_region
misc_feature
repeat_region
LTR
repeat_region
repeat_region
misc_feature
misc_feature
repeat_region
repeat_region
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
repeat_region
repeat_region
repeat_region
repeat_region
```

```
/rpt_family="Alu"
complement(25515. .25599)
/note="MALR00365"
/rpt_family="THE1"
complement(25777. .25901)
/note="MALR00322"
/rpt_family="MStrc"
complement(26808. .27050)
/note="X66298"
/rpt_family="Alu"
complement(27009. .27125)
/note="HSAL004729"
/rpt_family="Alu"
complement(27361. .27462)
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(27482. .27721)
/note="GRAIL 2 good exon ; ORF; frame 2"
30922. 31155
/rpt_family="MER33"
complement(31197. .31991)
/note="GRAIL 2 good exon ; ORF; frame 1"
complement(31907. .34394)
/gene="VIP2R"
/gene="VIP2R"
complement(join(<31907. .32006,34182. .34394))
/product="vasoactive intestinal polypeptide receptor 2
precursor"
complement(31907. .32006)
/gene="VIP2R"
/note="99% identity -- L40764"
complement(join(<31908. .32006,34182. .34234))
/gene="VIP2R"
/codon_start=3
/product="vasoactive intestinal polypeptide receptor 2
precursor"
/protein_id="MAC09171.1"
/db_xref="GI:3004859"
/translacion="MTLLPALLTCWLIAPVNSIHPCRFHLETOEETKAEILLRS
QTERKH"
complement(33411. .33597)
/gene="VIP2R"
/note="100% identity -- 259087"
complement(34182. .34232)
/gene="VIP2R"
/note="100% identity -- L40764"
complement(34200. .34394)
/gene="VIP2R"
/note="92% identity -- L36566"
34432. 34694
/rpt_family="TAR1"
35000. .35058
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(35325. .35445)
/rpt_family="MER42"
complement(36238. .36296)
/note="HSAL004342"
/rpt_family="Alu"
complement(37248. .37455)
/note="MER4_00001"
/rpt_family="MER4"
complement(37454. .37580)
/note="84% identity"
/db_xref="dbEST:T62489"
complement(37828. .38086)
/note="270040"
/rpt_family="MER4"
complement(41427. .41642)
/rpt_family="MLT2B2"
complement(42308. .42622)
/rpt_family="MLT2B2"
43022. .43522
/rpt_family="Tigger2"
43881. .44000
```

```
misc_feature
44984. .45116
/note="GRAIL 2 good exon ; ORF; frame 2"
misc_feature
complement(46422. .46478)
/note="GRAIL 2 good exon ; ORF; frame 0"
complement(46801. .47121)
/rpt_family="Alu"
complement(47511. .47779)
/note="GRAIL 2 good exon ; ORF; frame 1"
complement(48123. .48196)
/rpt_family="MLT2"
complement(49320. .49497)
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(49934. .50215)
/note="GRAIL 2 good exon ; ORF; frame 2"
complement(51833. .51901)
/note="GRAIL 2 good exon ; ORF; frame 2"
complement(52214. .52500)
/note="268162"
/rpt_family="Alu"
54960. .55032
/rpt_family="LIMC2"
55462. .55537
/rpt_family="LIME2"
57335. .58234
/rpt_family="LI"
complement(58350. .58466)
/rpt_family="MER44"
58503. .58646
/rpt_family="LIMA9"
complement(58757. .59020)
/note="GRAIL 2 good exon ; ORF; frame 0"
62145. .62345
/note="GRAIL 2 good exon ; ORF; frame 2"
complement(62518. .64744)
/rpt_family="Tigger1"
complement(66883. .67070)
/note="GRAIL 2 excellent exon ; ORF; frame 0"
complement(67293. .67336)
/note="GRAIL 2 good exon ; ORF; frame 1"
67421. .67746
/rpt_family="Alu"
complement(68255. .68320)
Query Match 55.1%; Score 2047.2; DB 9; Length 225432;
Best Local Similarity 75.0%; Pred. No 0;
Matches 3236; Conservative 0; Mismatches 403; Indels 677; Gaps 27;
QY 20 GAAGGACGAGAGGGCGTTCCTCTTTGGACATTTTTCATCCCTCGTTTTTTTCAG 79
Db 120520 GGAGGACGAGCGGGCGTTCCTCGCTGGGCTTCTCATGCCAGCTTTTCTCCAG 120579
QY 80 ATGTGGCTTGGTCTGGCGCAAGTCCACGACGACGCTTAAGCTTACTTCTGTGAAA 139
Db 120580 GTGTGACTTGGTC--GGACGCGGATCCCGGACGCCAGCTGGGCTTAGTCTGTGAAA 120637
QY 140 GGGGAAAGTATCCCTGTGTGAAAGCGTTAACTTGTGGAGGGGTGGCGGACGTGAGTT 199
Db 120638 GAGTAAGTGCTCTCTCGGAGAGCGGGTAACCTCCCAAGGCGCTCGGAGCGGAGTT 120697
QY 200 CTTCCCATGCCAGCGAATGGTGTGGCTTTGAGCTGTGTCAGGAGCGGCGTTCGACGTGT 259
Db 120698 C-TCCCGAGCGGAGCGAATGGGCGGCGCTGAGCTGGTCTA-GAGCGGCTCCACGCGT 120755
QY 260 CTGAGGAGCGGCGGAGGGGGGAGGTGGCCACACAGACCGCGGTCTCTTAAGAGA 319
Db 120756 CTGAGGAGCGGCGGCGGCGGAGGTGGCCAGAGTGGCCAGAG--CGCGGCTTCTGTGAAGGA 120812
QY 320 CGTTGGGAAG---ATTTCGATTTCGAGAAAGAAACCGGATTGAAAGAGCCAGGCC 376
Db 120813 CTGAGGAGGAGAGATTCCATTCGACAGAGGAGGAGCCGATTCAGAG--CCAGGCC 120871
```


Db 123006 TCCCTTCTCATGAAGGGCATAGTGCTCAGACAGTTGAGGATGCGAGCAACCCCTC 123005
QY 2285 CACTCCAGAGTTGGGTGACGAGATTTCAACTGGCCACCCCTTTGCCCTCCACTATTGAA 2344
Db 123066 CACTCCAGAATTGGGTGG-ATGGATTTTCAATGGGCTAAC--TTGGCTTTCACAAATGA 123122
QY 2345 TTTTTCAGACCCCATTTCTCATGCTGGAATGGATGCTGGAGCTTGGCAGCTTTCT 2404
Db 123123 GCTTTTTCAGACCCCATTTCTCATGCTGGAATGGATGCTGGAGCTTGGCAGTTTCT 123182
QY 2405 TTTCCCTCTGCTTTGACTAGGAACCGAGCTCTTAATTTCTCAGGACAGACTAGCTGGCA 2464
Db 123183 TTTCCCTCATTTTCACAGGAGCTGGACTATAATTTCTCAGGACAGACTAGCTGGCA 123242
QY 2465 CATTAATCCCTACCTTAGTTCTTTCTCTGACTCTCTGGAAGATACTCTCTGTAATCTCTG 2524
Db 123243 CATTAATCCACCTCAGTTCTTCTCTCTGATCCCTTGGGAAGATCCCTGTAATCTCTG 123302
QY 2525 TAAAGGTTTTGGGGATTAAGGTGTTTAAACCACTCCAGCTTCTCTCTCTTTTCTTTT 2584
Db 123303 TAAAGGTTTTGGGGATTAAGGTGTTTAAACCACTCCCAACTTTCTCTTTTATTTCT- 123361
QY 2585 TTTCTGAAAAAGGAAAAAGCACACAGCACACAATTTCAAGCCATTTTCAGATCAGAACTC 2644
Db 123362 ---GAATAAGGAAAAATGGCCACACAGCAATTTCAAGCCAGTTTCAGATCAAAACCC 123418
QY 2645 CAGAAGTGTGACAGATGCTTATTCGTAGAGTTCCCTCAGAGAGCCATGCTGTTTATG 2704
Db 123419 CAAAAGTGTGAGGAGATGCTTAGTCATAAGATTCCCTCAGAGAACCATGGTGTGTTG 123478
QY 2705 AAGAGAAGAGTAGTATGCTCTGCCAGAGAGAGCTCTCTTTAACTCTCTCTCTCTCT 2764
Db 123479 AAGAGAAGTGTGAGTATGCTCTGCCAGGA-CAGCTCTCTTTAACTCTCTCTCTCTCT 123537
QY 2765 ATGAATTTCTTAAGGCTGAAGGAATCAAGAGAGTGGGATGGAATCTTTATCCCTT 2824
Db 123538 ATGAATTTGTTAAGGCTGGAGGATGGAGAGTGGGATGGAATCTTTTACCCCTT 123597
QY 2825 TGTGTTAAACAGAGGAGCCATGAGGCTGGGAGATCATAGCCCTTCTAGGCAAGATCTCT 2884
Db 123598 TTTTAAACATGGGCGAGCCGTGGGCTGGGAGATCATAGCCCTTCTAGGCAAGATCTCT 123657
QY 2885 GTTCACTGCCAGCTATAGTAATTTACTATTTTCAATTTCAATATATTTCTGTTCT 2944
Db 123658 ACTCACTGCCAGGCTAATGATTATTTACTGTTTTCGAATTTGAAATATATTC-GGTTG 123716
QY 2945 TTTTCTAAATGTGAAGACTTACCAATGAATTTTAGATCATTTCCAGAGGAGATTTT 3004
Db 123717 TTTTGAATATGAACCTATCAATGAATTTTAGATCATTTCCAGAGGAGATTTCTT 123776
QY 3005 TTGCTCTCTCATCTTTTCCACAGAGTGTCTCTGTTTGTGAGCTTAAGGTAAGAGGG 3064
Db 123777 TT-CTCTTCCCATCTTTTCCACAGTGTCTCTCTGTTTGTGAGCTTAAGGTA----GAG 123830
QY 3065 ACATCTCTGCTGTTTAAACAGAGCTCCATATCTGTGAGGCCAGCA-- 3111
Db 123831 ACACCTCTGCTGTTTAAACAGGAGCTCCATATCTATGAGGCTAGCAAAATTTTCTTTTC 123890
QY 3112 ----- 3111
Db 123891 TTTTCTTTTGAACAGCTCTCACACTGTGCGCTGGCTAGAGTGCAGTGGCGGATCTTG 123950
QY 3112 ----- 3111
Db 123951 GCTCACTGCAACCTCCACCTCTCTGGGTTCAAGCGATTCTCTGCTCAGCTCCCAAGTA 124010
QY 3112 ----- 3111
Db 124011 GCTGGATTACAGGTCCTCGCCACAGGCGAGCTAAATTTTGTGTTTTTAGTAGAGAT 124070
QY 3112 ----- 3111

Db 124071 GGGGTTTCACTATGTAGCCAGGCTGGTCTCAACTCCTGACCTCGTGTGATCCGCCACCTC 124130
QY 3112 -----ATATTTTC 3119
Db 124131 GGCCCCCAAAGTGTGGGTTACAGCGTGAAGCCACCGCCCGCCACAGATATTTTC 124190
QY 3120 TTAACATCATGGGACACAGCAGATTTCTGCCCTGGTGGAGGTGCTGTGTCATATGT 3179
Db 124191 TTAACATCATGAGAGAGACAGATTTCTGCCCTCAGTGAAGTCATGCTGTGCTGTGT 124250
QY 3180 CCTACCCCTCTGCTTCATGTCAGGGAAGTTGGAATGGGGGCTACATATATGCTCTCTCT 3239
Db 124251 CCTA-CCCGCTGCTTCATGTCAGGAGGTGGAAATGGGGCGCATGTGCCCTCTCT- 124308
QY 3240 CCGCTACAAAGATGTGTGTTTCCATCTGATCTTCCACTCTTGTGAGGGGAAGAAG 3299
Db 124309 CCCATCTACAAGAGTTGTGTGTTTCCCATCTGATCTTCCACTCTTGTAGGGGAGGAAG 124368
QY 3300 GGGCCTGGTATCTCAGGAGATTTGTTGAATCTGCTGTCTATCCCTTCTATCCCACT 3359
Db 124369 GGGCCTGGTATCTCAGGAGATTTGTTGAATCTGCTGTCTATCCCTTCT- 124424
QY 3360 GCCTTCATATATGTTAGCCCATACCCCAATAACTGCTATATTAGACACCCCGAGCCA 3419
Db 124425 ACCTGATAGTAGTTAGCCCATACCCCAATAACTGCTATATTAGACACCCCGAGCCA 124484
QY 3420 GTTTCGGCTGCTGCTGTCT-TTGCTGCGCATGTTTTCACAGAGGAAG- AATTTCTTGTCT 3477
Db 124485 GTTTCGGCTGCTGCTGTGCGCATGTTTTCACAGAGGAAGAAACAATAT 124544
QY 3478 ATTTTCTTTCATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3537
Db 124545 GCTATTTTTCATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 124604
QY 3538 CTGTTAGGGGTGGGAGGAATATTTCAGGAGGCGTGGGTCTTTAGGAAAGGAATGGGA 3597
Db 124605 CTGTAATGGGTGGGAGGAATATTTCAGGAGGCGTGGGTCTTTAGGAAAGGAATGGGA 124664
QY 3598 AGCAACATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3657
Db 124665 ATCAACATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 124724
QY 3658 ATCAATATAAAGTATATATGTTTAAATGAATTAATTAATTAATTAATTAATTAATTA 3713
Db 124725 ATGCAATATAAAGTATATGTTTAAATGAATTAATTAATTAATTAATTAATTAATTA 124780

RESULT 8
HSU78576
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES

HSU78576 2133 bp mRNA linear PRI 20-DEC-1996
Human 68 kDa type I phosphatidylinositol-4-phosphate 5-kinase alpha
mRNA, clone PIP5K1a2, complete cds.
U78576
U78576.1 GI:1743872
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2133)
Loijens, J.C. and Anderson, R.A.
Type I phosphatidylinositol-4-phosphate 5-kinases are distinct
members of this novel lipid kinase family
J. Biol. Chem. 271 (51), 32937-32943 (1996)
97115834
8955136
2 (bases 1 to 2133)
Loijens, J.C. and Anderson, R.A.
Direct Submission
Submitted (18-NOV-1996) Pharmacology, University of Wisconsin -
Madison, 1300 University Ave., Madison, WI 53706, USA
Location/Qualifiers
1. .2133

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="pip5k1a2"
/clone_lib="human fetal brain 5'-stretch cDNA library
(Clontech # HL1149x)"
422..2110
/EC_number="2.7.1.68"
/notes="isoforms, possibly alternatively spliced, encoded
by GenBank Accession Numbers U78575 and U78577"
/codon_start=1
/product="68 kDa type I phosphatidylinositol-4-phosphate
5-kinase alpha"
/protein_id="AAC50911.1"
/db_xref="GI:1743873"
/translation="MASASSGSPSSSVFSSFDPAVPSCITLSSASGKIKPMASEVLEA
KQSYISLVPYASGMPIKIKHRSVDSGETTKYKTKSSALKGAIQIGITHTVGSLSST
RPERVMDIFFYVESIFPPSEGSNLTAPAHYNDRFKTYAPVAFRYFRELFIGRPDD
YIYSLSEPIECSGASGLSEYSSDDEFIITQVHKREAEFLQKLLPGLYWNKLNDD
PRLPKFYGLYQVAGKNIRIVVNNLLPRSVKMHKIDYDLKGSYKRRASQREK
PLPFLDLOQIDPDLADMYNAICTLQDCLVLQSEKIMDYSLIMS1HNIDH
AOREPLSSTQYSDVTRRPAPOKALYSTAMESIQEARRGSMETDHDHMGGIPARNSK
GERLLXIGIIDLQYRFVKLEHSWKALVHOGDVSVHRPGFARFRQFPCNTVF
KKIPLKPSKRSKRSFSSRRAGSSNSCITYOPSVSGHKAQVTKAEVPEGVHLG
RPDLQPTPELEISGSPIDPFSFPLVGETLQMLTSTTLEKLEVAESEFTH"
2110..>2133
3'UTR
BASE COUNT 524 a 524 c 559 g 526 t
ORIGIN

Query Match 54.2%; Score 2014; DB 9; Length 2133;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 2073; Conservative 0; Mismatches 0; Indels 39; Gaps 2;

QY 1 ATTAACAGCGCTGTAGGAAGACGACGAGAGGGCGTTCGCTCCCTTGGGACCTTTCA 60
DB 22 ATTAACAGCGCTGTAGGAAGACGACGAGAGGGCGTTCGCTCCCTTGGGACCTTTCA 81
QY 61 TGCCTCGCTTTTTCAGATGTGCTGTGGGCGCAAGGTCCAGAGCCAGCTTA 120
DB 82 TGCCTCGCTTTTTCAGATGTGCTGTGGGCGCAAGGTCCAGAGCCAGCTTA 141
QY 121 AGCTTACTCTCTGAAAGGGGAAAGTATCCCTCTGTGGAAGCGGTAACTTGTGGAG 180
DB 142 AGCTTACTCTCTGAAAGGGGAAAGTATCCCTCTGTGGAAGCGGTAACTTGTGGAG 201
QY 181 GGGTGGCGGACGTGAGTCTTCCCATGCCAGCGAATGGTGTGGCTTCAGCTGTGTC 240
DB 202 GGGTGGCGGACGTGAGTCTTCCCATGCCAGCGAATGGTGTGGCTTCAGCTGTGTC 261
QY 241 AGGAGCGGCTCGACGTGTCTGAGGAGGCGCGGAGGGCGGAGGTGGCCACACAGAA 300
DB 262 AGGAGCGGCTCGACGTGTCTGAGGAGGCGCGGAGGGCGGAGGTGGCCACACAGAA 321
QY 301 CGCGGGTCTCTGAAGAGAGCTTGGGAAGATTCGATCCGAGAGAGAGAAACCGGATT 360
DB 322 CGCGGGTCTCTGAAGAGAGCTTGGGAAGATTCGATCCGAGAGAGAGAAACCGGATT 381
QY 361 GAAAGAGAGCCAGCCGCTGAGGGGAGGGGCTGCTAAGATGCGTCCGCTCCTCGG 420
DB 382 GAAAGAGAGCCAGCCGCTGAGGGGAGGGGCTGCTAAGATGCGTCCGCTCCTCGG 441
QY 421 GCCGTCTCTTGGTCTGGTTCATCTTGTATCCGCGGCTCCCTTCCCTGCTACCTGTGC 480
DB 442 GCCGTCTCTTGGTCTGGTTCATCTTGTATCCGCGGCTCCCTTCCCTGCTACCTGTGC 501
QY 481 CT----CAGCATCTGAATCAAGAGACCCATGGCATCTGA----- 516
DB 502 CTCAGCAGCATCTGGAATCAAGAGACCCATGGCATCTGAGGTCTTGGAAAGCTAGACAGGA 561
QY 517 -----GCTGCTTATGCTCTGGCATCCCATCAAGAAATAGGCCATAG 561
DB 562 TTCTTACATCTCAATGTGTGCTTATGCTCTGGCATCCCATCAAGAAATAGGCCATAG 621
QY 562 AAGTGTGATTCTCTCAGGAGAGACAACATATAAAGACACACCTCATCAGCCTTGAAGG 621

```

```

|||||
622 AAGTGTGATTCTTCAGAGAGACACATATAAAGACACCTCATCAGCTTGAAGG 681
QY 622 TGCCATCCAGTTAGGCATATCCACACTGTGGGAGCCTGAGTACCAACACGAGCGTGA 681
DB 682 TGCCATCCAGTTAGGCATATCCACACTGTGGGAGCCTGAGTACCAACACGAGCGTGA 741
QY 682 TGTCTCATGCAAGATTCTACGTGGTGTGAGAGTATCTTCTTCCAGTGAAGGAGCAA 741
DB 742 TGTCTCATGCAAGATTCTACGTGGTGTGAGAGTATCTTCTTCCAGTGAAGGAGCAA 801
QY 742 CTGACCCCTGCTCATCACTACAATGACTTTCGTTTCAAGACCTATGACCTGTTCCTT 801
DB 802 CTGACCCCTGCTCATCACTACAATGACTTTCGTTTCAAGACCTATGACCTGTTCCTT 861
QY 802 CCGCTACTTCCGGGAGCTATTGTTGATATCCGCGCGGATGATTACTTGTATTCCTCTGCAG 861
DB 862 CCGCTACTTCCGGGAGCTATTGTTGATATCCGCGCGGATGATTACTTGTATTCCTCTGCAG 921
QY 862 TGAGCCGCTGATTGAACCTCTGAGCTCTGGAGCTAGTGGTCCCTTATCTATGTCTCCAG 921
DB 922 TGAGCCGCTGATTGAACCTCTGAGCTCTGGAGCTAGTGGTCCCTTATCTATGTCTCCAG 981
QY 922 CGAGCATGAGTTCATTATTAAGACAGTCCAACATAAAGAGCGGAATTTCTGCAAGACT 981
DB 982 CGAGCATGAGTTCATTATTAAGACAGTCCAACATAAAGAGCGGAATTTCTGCAAGACT 1041
QY 982 GTTCTCAGGATPACTACATGAACCTCAACAGAACCTCGGACTTTGCTGCCTTAATTTCTA 1041
DB 1042 GTTCTCAGGATPACTACATGAACCTCAACAGAACCTCGGACTTTGCTGCCTTAATTTCTA 1101
QY 1042 TGGACTGTACTGTGTGAGGAGCTGGCAAGACATTCGGATTGTGGTGTATGAACAATCT 1101
DB 1102 TGGACTGTACTGTGTGAGGAGCTGGCAAGACATTCGGATTGTGGTGTATGAACAATCT 1161
QY 1102 TTTTACCAGATCGGTAAATAATGCATATCAAAATATGACCTCAAAAGCTCAACCTCAAAAG 1161
DB 1162 TTTTACCAGATCGGTAAATAATGCATATCAAAATATGACCTCAAAAGCTCAACCTCAAAAG 1221
QY 1162 CGGGCTTCCAGAAAGAGCGAGAGAGCCCTTCTCCACATTTTAAAGACCTAGACTTCTT 1221
DB 1222 CGGGCTTCCAGAAAGAGCGAGAGAGCCCTTCTCCACATTTTAAAGACCTAGACTTCTT 1281
QY 1222 ACAAGACATCCCTGATGGTCTTTTTTGGATGCTGACATGTACAGGCTCTCTGTAAGAC 1281
DB 1282 ACAAGACATCCCTGATGGTCTTTTTTGGATGCTGACATGTACAGGCTCTCTGTAAGAC 1341
QY 1282 CCTGACGGTCTGACTGTTTGGTGTGACAGCTTCAAGATAATGGATTACAGCTCTTGTAT 1341
DB 1342 CCTGACGGTCTGACTGTTTGGTGTGACAGCTTCAAGATAATGGATTACAGCTCTTGTAT 1401
QY 1342 GTCAATCCATATATAGATCATGCACACGAGAGCCCTTAAAGCAGTGAACACAGTACTC 1401
DB 1402 GTCAATCCATATATAGATCATGCACACGAGAGCCCTTAAAGCAGTGAACACAGTACTC 1461
QY 1402 AGTTGATCTCGAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1461
DB 1462 AGTTGATCTCGAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1521
QY 1462 CCAGGAGAGGCTCGACGGGCTGTTGATCCATGAGAGCTGATGACCATATGGGTGGCATCCC 1521
DB 1522 CCAGGAGAGGCTCGACGGGCTGTTGATCCATGAGAGCTGATGACCATATGGGTGGCATCCC 1581
QY 1522 TGCCCGGAATAGTAAAGGGGAAAGGCTTCTGCTTTATATTGGCATCATTTGACATTTCTACA 1581
DB 1582 TGCCCGGAATAGTAAAGGGGAAAGGCTTCTGCTTTATATTGGCATCATTTGACATTTCTACA 1641
QY 1582 GTCCTACAGGTTTGTAAAGAGTGGAGCCTCTTGGAAAGCCCTGGTGTACATGACGGAGA 1641
DB 1642 GTCCTACAGGTTTGTAAAGAGTGGAGCCTCTTGGAAAGCCCTGGTGTACATGACGGAGA 1701
QY 1642 CACTGTCTCAGTGCATGCCCGAGGCTTCTAGCTGAAAGGTTCCAGCGCTTCATCTGCA 1701
|||||

```

CDS

Db	842	TTACTTGTATTCCTCTCGCAGTGCAGCGCTGATTGAACCTCTGTAGCTCTGGAGCTAGTGG	901
Qy	901	TTCCCTATTCTATGTCTCCAGCGAGATGAGTTCAATTATTAAGACAGCTCCAACATAAAGA	960
Db	902	TTCCCTATTCTATGTCTCCAGCGAGATGAGTTCAATTATTAAGACAGCTCCAACATAAAGA	961
Qy	961	GGCGGAATTTCTGCAGAAAGCTGCTTTCCAGGATACACATGAAACCTCAACAGAACCCCTCG	1020
Db	962	GGCGGAATTTCTGCAGAAAGCTGCTTTCCAGGATACACATGAAACCTCAACAGAACCCCTCG	1021
Qy	1021	GACTTTGCTGCCTTAAATCTATGGAAGTGTACCTGTGTCGAGCGAGTGGCAAGAACATTCG	1080
Db	1022	GACTTTGCTGCCTTAAATCTATGGAAGTGTACCTGTGTCGAGCGAGTGGCAAGAACATTCG	1081
Qy	1081	GATTGGTGTGATGAACAATCTTTTACCAGAGTCGGTAAAAATGCATATCAAAATATGACCT	1140
Db	1082	GATTGGTGTGATGAACAATCTTTTACCAGAGTCGGTAAAAATGCATATCAAAATATGACCT	1141
Qy	1141	CAAAGGCTCAACTACAAAGCGGGGCTTCCAGAAAAGAGCGAGAGAAGCCCTTTCCAC	1200
Db	1142	CAAAGGCTCAACTACAAAGCGGGGCTTCCAGAAAAGAGCGAGAGAAGCCCTTTCCAC	1201
Qy	1201	ATTTAAAGACCTAGACTCTCTTACAAGACATCCCTGATGCTCTTTTGTGGATGCTGACAT	1260
Db	1202	ATTTAAAGACCTAGACTCTCTTACAAGACATCCCTGATGCTCTTTTGTGGATGCTGACAT	1261
Qy	1261	GTACAAAGCTCTCTGTAAAGACCCCTGCAGCGTGACTGTTGGTGTGCAGAGCTTCAAGAT	1320
Db	1262	GTACAAAGCTCTCTGTAAAGACCCCTGCAGCGTGACTGTTGGTGTGCAGAGCTTCAAGAT	1321
Qy	1321	AATGGATTAACAGCCTCTTGATGTCAATCCATAAATATAGATCATGCACAACGAGAGCCCTT	1380
Db	1322	AATGGATTAACAGCCTCTTGATGTCAATCCATAAATATAGATCATGCACAACGAGAGCCCTT	1381
Qy	1381	AAGCAGTGAACACAGTACTCAGTTGATCTCGAAGACCGGCCCCCCCCAAAAGGCTCTGTA	1440
Db	1382	AAGCAGTGAACACAGTACTCAGTTGATCTCGAAGACCGGCCCCCCCCAAAAGGCTCTGTA	1441
Qy	1441	TTCCACAGCCATGGAAATCCATCCAGGAGAGGCTCGACGGGGTGGTACCATGAGAGCTGA	1500
Db	1442	TTCCACAGCCATGGAAATCCATCCAGGAGAGGCTCGACGGGGTGGTACCATGAGAGCTGA	1501
Qy	1501	TGACCATATGGGTGGCATCCCTTGCCCGGAATAGTAAGAGGGAAGGCTTTCGCTTTATAT	1560
Db	1502	TGACCATATGGGTGGCATCCCTTGCCCGGAATAGTAAGAGGGAAGGCTTTCGCTTTATAT	1561
Qy	1561	TGCAATCATGACATTTACAGTCTTACAGGTTTCTTAAGAAGTTGGAGCACTCTTGAA	1620
Db	1562	TGCAATCATGACATTTACAGTCTTACAGGTTTCTTAAGAAGTTGGAGCACTCTTGAA	1621
Qy	1621	AGCCCTGGTACATGACGGAGACACTGTCTCAGTGATCGCCAGGCTTCTACGCTGAACG	1680
Db	1622	AGCCCTGGTACATGACGGAGACACTGTCTCAGTGATCGCCAGGCTTCTACGCTGAACG	1681
Qy	1681	GTTCCAGCGCTTCAATGTGCAACACAGTATTTAAGAAGATTCCCTTTGAAGCCCTTCCCTTC	1740
Db	1682	GTTCCAGCGCTTCAATGTGCAACACAGTATTTAAGAAGATTCCCTTTGAAGCCCTTCCCTTC	1725
Qy	1741	CAAAAAGTTTCGGTCTGGCTCATCTTCTCTCGGCGAGCGGCTCCAGTGGCAACTCCGT	1800
Db	1726	-----	1725
Qy	1801	CATTACTTACCAGCCATCGGTCTCTGGGAGACAAAGGCAACAAGTGACAAAGGCGAGA	1860
Db	1726	-----	1725
Qy	1861	AGTGGAGCCAGGCGTTTCACTTTGGTCGCTGATGTTTAACTTCAGACTCCACCCCTTTGGA	1920
Db	1726	-----GGGTTCACCTTGGTCGCTGATGTTTAACTTCAGACTCCACCCCTTTGGA	1774
Qy	1921	GGAAATCAGTAGGGGCTCGCCTATTCTTGACCCCAAGTTTCTACCTCTAGTTGGAGAGAC	1980
Db	1775	GGAAATCAGTAGGGGCTCGCCTATTCTTGACCCCAAGTTTCTACCTCTAGTTGGAGAGAC	1834

Qy	1981	TTTGAAATGCTAATCAAGTACAACCTTGGAAAAGCTTGAAGTTGCAGAGTCAGAGTT	2040
Db	1835	TTTTGAATGCTAACTACAAGTACAACCTTGGAAAAGCTTGAAGTTGCAGAGTCAGAGTT	1894
Qy	2041	CACCCATTAAAGCGCAAAGCCTCAGAAGACCTTGGAAACAAGATCTGCCATCTCTGTGATCC	2100
Db	1895	CACCCATTAAAGCGCAAAGCCTCAGAAGACCTTGGAAACAAGATCTGCCATCTCTGTGATCC	1954
Qy	2101	CAAGATGTCAGCCCTTGCCCCAGCAATGCTGAATTTCTTACTTGGTCATCAAA	2156
Db	1955	CAAGATGTCAGCCCTTGCCCCAGCAATGCTGAATTTCTTACTTGGTCATCAAA	2010

RESULT 10					
AL365510/c					
LOCUS	AL365510	180189 bp	DNA	linear	PRI 08-JUN-2001
DEFINITION	Human DNA sequence from clone RP11-429H9 on chromosome 10, complete sequence.				

SOURCE	HOMO sapiens human.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE	1 (bases 1 to 180189)

JOURNAL
Direct submission
12185
Submitted (08-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk

On June 11, 2001 this sequence version replaced g1:14272312. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/nor/chr11>
 RP11-429g9 is from the library RP11-11.2 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-429H9. The true left end of clone RP11-162K11 is at 109726 in this sequence. The true right end of clone RP11-397J18 is at 42873 in this sequence.

```

FEATURES
  Location/Qualifiers
    source
      1..180189
true right end of clone RP11-39/J16 is at 426/3 in this sequence.

```

```
repeat_region
1. .224
/cione_1112- rfc1_111.2
/note="ML2FB repeat: matches 1. .227 of consensus"
236. .382
repeat_region
/note="LIMC4 repeat: matches 7618. .7776 of consensus"
```



```

misc_feature      complement(495..1074)
/note="match: STS: Em:HS1100C7"
2802..3002
/note="MIR repeat: matches 46..244 of consensus"
4688..5178
/note="L1P13 repeat: matches 5648..6144 of consensus"
5188..5501
/note="AluX repeat: matches 1..312 of consensus"
5959..6225
/note="AluX repeat: matches 2..284 of consensus"
6739..6839
/note="MIR repeat: matches 80..191 of consensus"
6905..7200
/note="AluB repeat: matches 3..299 of consensus"
7268..8052
/note="L1A10 repeat: matches 5502..6287 of consensus"
8319..8491
/note="FRAM repeat: matches 0..173 of consensus"
9136..9446
/note="AluY repeat: matches 1..311 of consensus"
9447..9731
/note="AluJ repeat: matches 1..291 of consensus"
10082..10259
/note="FRAM repeat: matches -3..169 of consensus"
10265..10585
/note="AluX repeat: matches 1..301 of consensus"
10667..10910
/note="MIR repeat: matches 1..241 of consensus"
12586..12885
/note="AluSc repeat: matches 1..300 of consensus"
13009..19121
/note="L1P15 repeat: matches 11..6144 of consensus"
19127..19218
/note="46 copies 2 mer aa 63% conserved"
19221..19287
/note="L12 repeat: matches 2574..2634 of consensus"
19341..19579
/note="L1B18 repeat: matches 5400..5638 of consensus"
19581..19923
/note="MER61A repeat: matches 5..354 of consensus"
19925..20445
/note="L1B18 repeat: matches 5640..6171 of consensus"
20641..20958
/note="AluX repeat: matches 1..298 of consensus"
21217..21766
/note="L1M4 repeat: matches 6342..6929 of consensus"
21772..21985
/note="AluB repeat: matches 86..297 of consensus"
21992..22122
/note="L1M4 repeat: matches 6934..7054 of consensus"
22123..22418
/note="AluX repeat: matches 1..301 of consensus"
22419..23122
/note="L1M4 repeat: matches 7054..7829 of consensus"
23182..23884
/note="L1M4 repeat: matches 6955..7618 of consensus"
23885..24166
/note="AluB repeat: matches 1..269 of consensus"
24167..24275
/note="L1M4 repeat: matches 7618..7730 of consensus"
24451..24542
/note="L12 repeat: matches 2624..2709 of consensus"
24755..25030
/note="MLT1A2 repeat: matches 6..313 of consensus"
25086..25135
/note="25 copies 2 mer tc 80% conserved"
25149..25458
/note="AluJ repeat: matches 1..296 of consensus"
25462..25535
/note="MLT1A2 repeat: matches 321..394 of consensus"
25943..26296
/note="MLT1A1 repeat: matches 1..343 of consensus"
29101..29479
repeat_region    /note="L2 repeat: matches 2156..2525 of consensus"
29481..30004
/note="L1M8 repeat: matches 5631..6153 of consensus"
30007..30210
/note="L2 repeat: matches 2531..2750 of consensus"
30338..30439
/note="MIR repeat: matches 57..165 of consensus"
30518..30576
/note="L2 repeat: matches 2690..2750 of consensus"
31508..31562
/note="L2 repeat: matches 2646..2704 of consensus"
31514..31717
/note="MIR repeat: matches 53..256 of consensus"
33399..33435
/note="tRNA-Leu-CTT repeat: matches 1..37 of consensus"
33482..33521
/note="MER5B repeat: matches 138..178 of consensus"
33523..33653
/note="FLAM_A repeat: matches 13..142 of consensus"
33673..33799
/note="MER5B repeat: matches 1..129 of consensus"
34148..35176
/note="L1M2 repeat: matches 4995..6089 of consensus"
35569..35673
/note="L2 repeat: matches 2634..2741 of consensus"
36553..36723
/note="AluSg/x repeat: matches 121..291 of consensus"
37781..38293
/note="MER21B repeat: matches 5..498 of consensus"
38317..38680
/note="L1P2 repeat: matches 5782..6146 of consensus"
38795..38944
/note="MER77 repeat: matches 462..791 of consensus"
39148..39249
/note="Tigger3b repeat: matches 1..102 of consensus"
39248..39571
/note="Tigger3b repeat: matches 274..584 of consensus"
39572..39928
/note="THE1B repeat: matches 1..364 of consensus"
39929..40526
/note="Tigger3b repeat: matches 584..1231 of consensus"
40527..40662
/note="MIR repeat: matches 16..154 of consensus"
41996..42091
/note="48 copies 2 mer tt 70% conserved"
42094..42372
/note="AluB repeat: matches 6..310 of consensus"
43374..43488
/note="MIR repeat: matches 138..262 of consensus"
43634..43785
/note="MER5B repeat: matches 2..156 of consensus"
43800..43895
/note="MER5B repeat: matches 2..130 of consensus"
43896..43979
/note="42 copies 2 mer tt 78% conserved"
/note="match: STS: Em:G34800"
47248..47367
/note="MER5A repeat: matches 50..169 of consensus"
47252..47394
/note="MER5A repeat: matches 36..183 of consensus"
47445..47516
/note="MSTB repeat: matches 1..73 of consensus"
47552..47619
/note="L2 repeat: matches 2641..2709 of consensus"
47628..47683
/note="MLT2B repeat: matches 394..448 of consensus"
47684..47815
/note="66 copies 2 mer ta 61% conserved"
47842..48211
Query Match      49.4%; Score 1835; DB 9; Length 180189;
Best Local Similarity 95.9%; Pred. No. 0;

```

Matches 1921: Conservative 0; Mismatches 65; Indels 17; Gaps 3;

Qy 1 ATTAACAGGCGTGGTTAGGAAGACAGGAGAGAGGGGCGTCCCTCTTTGGGACTTTTCA 60
 Db 47206 ATTAACAGGCGTGGTTAGGAAGACAGGAGAGAGGGGCGTCCCTCTTTGGGACTTTTCA 47147
 Qy 61 TGCCTCG-TTTTCTTTTTCAGATGTGGTCTGTGGCGCAAGGTCCAGCAGCAGCTT 119
 Db 47146 TGCCTCGTTTCTTTTTCAGATGTGGTCTGTGGCGCAAGGTCCAGCAGCAGCTT 47087
 Qy 120 AAGCTTACTCTCTGTGAAGGGGAAAGATATCCCTGTGTGAAGCGGTAAACTTTGTGA 179
 Db 47086 AAGCTTACTCTCTGTGAAGGGGAAAGATATCCCTGTGTGAAGCGGTAAACTTTGTGA 47027
 Qy 180 GGGGGTGGGACGTAGTCTTCCCATGCCAGCGCAATGGTGGCTTTAGCTGGTC 239
 Db 47026 GGGGGTGGGACGTAGTCTTCCCATGCCAGCGCAATGGTGGCTTTAGCTGGTC 46967
 Qy 240 CAGGAGCGGCTCGACGTGTCTGAGGAGGCGCGAGGGGCGGGAGGTGGCCACAGA 299
 Db 46966 CAGGAGCGGCTCGACGTGTCTGAGGAGGCGCGAGGGGCGGGAGGTGGCCACAGA 46907
 Qy 300 ACGCGGTTCTGTAAAGAGAGCTTTGGGAAGATTCGATTCGAGAGAGAGAAACCGGAT 359
 Db 46906 ACGTGGTCTCTATAAGAGACATTTGGGAAGATTCGATTCGAGAGAGAGAAACCGGAT 46847
 Qy 360 TGAAGAGAGCGGCGCTCAGGGGAGGGGCTGTAAAGTGGCTGGCCCTCTCCG 419
 Db 46846 TGAAGAGAGCGGCGCTTGAAGGGGAGGGGCTGTAAAGTGGCTGGCCCTCTCCG 46787
 Qy 420 GCGCTCTCTCTCGTGGTGTCTTTCATCTTTGTATCCCGGCTCCCTTCCTGTACCTTGT 479
 Db 46786 GCGCTGTCTCTCGTGGTGTCTTTCATCTTTGTATCCCGGCTCCCTTCCTGTACCTTGT 46728
 Qy 480 CCTCAGCATCTGGAATCAAGAGACCCATGGCATCTGAGGTGCCCTTATGCGCTCTGGCATGC 539
 Db 46727 CCTCAGCATCTGGAATCAAGAGTCCCATGGCATCTGAGGTGCCCTTATGCGCTCTGGCATGC 46668
 Qy 540 CCATCAAGAAAATAGGCCATAGAAGTGTGATTCCTCAGGAGAGACAAATATAAAGA 599
 Db 46667 CCATCAAGAAAATAGGCCATAGAAGTGTGATTCCTCAGGAG-----GGA 46623
 Qy 600 CAACCTCATAGCCTTGAAGGTGCCATCCAGTTAGGCATTACCCACACTGTGGGAGCC 659
 Db 46622 CAACCTCATAGCCTTGAAGGTGCCATCCAGTTAGGCATTACCCACACTGTGGGAGCC 46563
 Qy 660 TGAGTACCAACACAGAGCGTGTCTCTCATGCAAGATTTCTAGCTGGTTGAGAGATCT 719
 Db 46562 TGAGTACCAACACAGAGTGTCTCTCATGCAAGATTTCTAGCTGGTTGAGAGATCT 46503
 Qy 720 TCTTTCCAGTGAAGGAGCAACCTGACCCCTGTCTCATCAATGACATGCTTTGTTTCA 779
 Db 46502 TCTTTCCAGTGAAGGAGCAACCTGACCCCTGTCTCATCAATGACATGCTTTGTTTCA 46443
 Qy 780 AGACTATGACCTTGTTCCTTCGCTTCCGCTACTTTCGCGAGCTATTGGTATCCGCGCGATG 839
 Db 46442 AGACTATGACCTTGTTCCTTCGCTTCCGCTACTTTCGCGAGCTATTGGTATCCGCGCGATG 46383
 Qy 840 ATTACTTGTATTCCTCTGCAAGTGGCGCTGATTGAACTCTGAGCTCTGAGCTAGTG 899
 Db 46382 ATTACTTGTATTCCTCTGCAAGTGGCGCTGATTGAACTCTGAGCTCTGAGCTAGTG 46323
 Qy 900 GTTCCCTATTCTATGTCTCAGCAGCAGTATGATTCATTTAAGACAGTCCAAATAGA 959
 Db 46322 GTTCCCTATTCTATGTCTCAGCAGCAGTATGATTCATTTAAGACAGTCCAAATAGA 46263
 Qy 960 AGGCGGAATTTCTCAGAGCTGTCTCCAGGATACATCACTCAACCTCAACGACCCCTC 1019
 Db 46262 AAGCGGAGTTTCTCAGAGCTGTCTCCAGGATACATCACTCAACCTCAACGACCCCTC 46203
 Qy 1020 GGACTTTGCTGCCTAAATTTCTATGACTGTACTGTGTGAGGAGGTGGCAAGAACATTC 1079
 Db 46202 GGACTTTGCTGCCTAAATTTCTATGACTGTACTGTGTGAGGAGGTGGCAAGAACATTC 46143

Qy 1080 GGATTGTGGTGATGAACAAATCTTTTACCAGATCGGTAAAATATGATATCAAAATATGACC 1139
 Db 46142 GGATTGTGGTGATGAACAAATCTTTTACCAGATCGGTAAAATATGATATCAAAATATGACC 46083
 Qy 1140 TCAAGAGCTCAACCTTACAAACGGCGGCTTCCCAAGAGAGGAGAGAGCTCTTCCCA 1199
 Db 46082 TCAAGAGCTCAACCTTACAGGCGGCGGCTTCCCAAGAGAGGAGAGAGCTCTTCCCA 46023
 Qy 1200 CATTTAAAGACCTAGACTTCTTACAAGACATCCCTGTATGGTCTTTTGGATGCTGACA 1259
 Db 46022 CATTTAAAGACCTAGACTTCTTACAAGACATCCCTGTATGGTCTTTTGGATGCTGAGC 45963
 Qy 1260 TGTACAACGCTCTCTGTAAAGACCTTGCAGCGTGAATGTTTGGTGTGAGAGCTTCAAGA 1319
 Db 45962 TGCACAACGCTCTCTGTAAAGACCTTGCAGCGTGAATGTTTGGTGTGAGAGCTTCAAGA 45903
 Qy 1320 TAATGGATTACAGCCTCTTGTATGTCATATCAATCAATATATAGATCATGCACAAACGAGAGCCCT 1379
 Db 45902 TAATGGATTATAGCCTCTTGTATGTCATATCAATCAATATATAGATCATGCACAAACGAGAGCCCT 45843
 Qy 1380 TAAGCAGTGAACACAGTACTACTAGTTGATCTCGAAGACCGGCGGCGGCGGCTCTGT 1439
 Db 45842 TAAGCAGTGAACACAACTACTAGTTGATCTCGAAGACCGGCGGCGGCGGCTCTGT 45783
 Qy 1440 ATTCCACAGCCATGGAATCCATCCAGGAGAGGCTCGAGGGGTGGTACCATGGAGACTG 1499
 Db 45782 ATTCCACAGCCATGGAATCCATCCAGGAGAGGCTCGAGGGGTGGTACCATGGAGACTG 45723
 Qy 1500 ATGACCATATGGTGGCATCCCTGCCGCAATAGTAAAGGGGAAAGGCTCTGCTTTATA 1559
 Db 45722 ATGACCATATGGTGGCATCCCTGCCGCAATAGTAAAGGGGAAAGGCTCTGCTTTATA 45663
 Qy 1560 TTGGCATATTGACATTTACAGTCTTACAGTCTTAAAGAAAGTTGGAGCACTCTTGA 1619
 Db 45662 TTGGCATATTGACATTTACAGTCTTAAAGAAAGTTGGAGCACTCTTGA 45603
 Qy 1620 AAGCCCTGGTACATGACGAGACACTGTCTAGTGCATCGGCCAGGCTTCTACGCTGAAC 1679
 Db 45602 AAGCCCTGATACATGACGAGACACTGTCTAGTGCATCGGCCAGGCTTCTATGCTGAAT 45543
 Qy 1680 GGTTCACGCTTTCATGTGCAACACAGTATTTAAGAAAGTTCCCTTGAGCCCTTCTCCTT 1739
 Db 45542 GGTTCACGCTTTCATGTGCAACACAGTATTTAAGAAAGTTCCCTTGAGCCCTTCTCCTT 45483
 Qy 1740 CCAAAAAGTTTCGGTCTGGCTCATCTTCTCGCGAGCAGGCTCCAGTGGCAACTCT 1799
 Db 45482 CCAAAAAGTTTCGGTCTGGCTCATCTTCTCGCGAGCAGGCTCCAGTGGCAACTCT 45423
 Qy 1800 GCATTACTTACAGCCATCGGTCTCTGGGAAACAAAGGACAAAGTGAACAAAGGCG 1859
 Db 45422 GCATTACTTACAGCCATTTGGTCTCTGGGAAACAAAGGACAAAGTGAACAAAGGCGG 45363
 Qy 1860 AAGTGGAGCCAGGCTTCCACTTGGTCTCTGATGTTTACCTCAGACTCCACCTTTGG 1919
 Db 45362 AAGTGGAGCCAGGCTTCCACTTGGTCTCTGATGTTTACCTCAGACTCCACCTTTGG 45303
 Qy 1920 AGGAAATCAGTGAAGGCTCGCTTATCTGACCCAGTCTTCTACCTCTAGTTGAGAGA 1979
 Db 45302 AGGAAATCAGTGAAGGCTCGCTTATCTGACCCAGTCTTCTACCTCTAGTTGAGAGA 45243
 Qy 1980 CTTTGCATAATGCTACTACAAGT 2002
 Db 45242 CTTTGCATAATGCTACTACAAGT 45220

RESULT 11
 AL592111
 LOCUS Homo sapiens chromosome 1 clone RP11-124I9, *** SEQUENCING IN
 DEFINITION
 PROGRESS *** 8 unordered pieces.
 AL592111
 AL592111.3 GI:14575471


```
Db 92485 GATGAATTTCTTAAGGCTGAAGGAATGAAGAGAGTGGGACATGGGTAATCTTTATCCCT 92544
Qy 2824 TTTGTTAAACAGAGGAGCCATGGCTGGGAGATCATAGCCCTTCTAGGCGAATCC 2893
Db 92545 TTTGTTAAACAGAGGAGCCATGGCTGGGAGATCATAGCCCTTCTAGGCGAATCC 92604
Qy 2884 TGTTCACCTGCAGGCTAGTAATTAATTAATTTTCAATTTTGAATATATCTGCTTG 2943
Db 92605 TGTTCACCTGCAGGCTAGTAATTAATTAATTTTCAATTTTGAATATATCTGCTTG 92664
Qy 2944 TTTTCTAAATGTGAAGACTTACCAATGAATTTTAGATCATCTCCAGAGGAGATTTT 3003
Db 92665 TTTTCTAAATGTGAAGACTTACCAATGAATTTTAGATCATCTCCAGAGGAGATTTT 92724
Qy 3004 TTTGCTCTTCTCATCTTTTCCACAGAGTGTCTCTGTTTGTGGAGCTAAGTAAGAGG 3063
Db 92725 TTTGCTCTTCTCATCTTTTCCACAGAGTGTCTCTGTTTGTGGAGCTAAGTAAGAGG 92784
Qy 3064 GACACTTCTGTCTTTAAACAGAGCTCCATATCTGTGAGGCCAGCAATATTTCTTAA 3123
Db 92785 GACACTTCTGTCTTTAAACAGAGCTCCATATCTGTGAGGCCAGCAATATTTCTTAA 92844
Qy 3124 ACTCATGGGAGACAGAGATTTCTGCTTGGTGAGTCAATTTGCTGCCATATGCTCTA 3183
Db 92845 ACTCATGGGAGACAGAGATTTCTGCTTGGTGAGTCAATTTGCTGCCATATGCTCTA 92904
Qy 3184 CCCCCCTGCTTCATGAGGAGAGTGGAAATGGGGGCTACATATGAGCCCTCTCTCCCG 3243
Db 92905 CCCCCCTGCTTCATGAGGAGAGTGGAAATGGGGGCTACATATGAGCCCTCTCTCCCG 92964
Qy 3244 TCTACAAGAGTGTGGTTCCTCATCTCCACTCTGTGTCAGGGAGAGAGGGG 3303
Db 92965 TCTACAAGAGTGTGGTTCCTCATCTCCACTCTGTGTCAGGGAGAGAGGGG 93024
Qy 3304 CTGCTATCTCAGGAGAGTGTGAATTCCTGTTCTATCCCTTCTCTATCCACCCCTGCT 3363
Db 93025 CTGCTATCTCAGGAGAGTGTGAATTCCTGTTCTATCCCTTCTCTATCCACCCCTGCT 93084
Qy 3364 TGATAATATGTAGCCCATACCCCAATACGTCTATATATAGACACCCCGAGGCTTT 3423
Db 93085 TGATAATATGTAGCCCATACCCCAATACGTCTATATATAGACACCCCGAGGCTTT 93144
Qy 3424 CTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3483
Db 93145 CTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 93204
Qy 3484 TTTTCAATATTTACTATTTATGATGATTTTAAAGTGTGTTTATTAAGGACAGATTTCTGTTA 3543
Db 93205 TTTTCAATATTTACTATTTATGATGATTTTAAAGTGTGTTTATTAAGGACAGATTTCTGTTA 93264
Qy 3544 GGGGTGGGAGGAAATATTGAGGAGGCGTGGGCTTTAGGAAAGGAATGGGGAAGCAAC 3603
Db 93265 GGGGTGGGAGGAAATATTGAGGAGGCGTGGGCTTTAGGAAAGGAATGGGGAAGCAAC 93324
Qy 3604 ATTTTATTAAGTGTACTATTTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3663
Db 93325 ATTTTATTAAGTGTACTATTTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 93384
Qy 3664 TATAAAGTGATATATGTTTAAATGTAATAAAGTGTGTTTAAAGTGTATTTA 3713
Db 93385 TATAAAGTGATATATGTTTAAATGTAATAAAGTGTGTTTAAAGTGTATTTA 93434

RESULT 12
AL592424
LOCUS Human DNA sequence from clone Rp11-68118 on chromosome 1, complete
DEFINITION AL592424
ACCESSION AL592424
VERSION AL592424.13
KEYWORDS GI:20068493
SOURCE HTG.
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Direct Submission
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
clonerequest@sanger.ac.uk
humquery@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:19699452.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
Rp11-68118 is from the library RpCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBac3.6.

FEATURES
Location/Qualifiers
1..159148
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="Rp11-68118"
/clone_lib="RpCI-11.1"
BASE COUNT 40758 a 37585 c 37589 g 43216 t
ORIGIN

Query Match 44.6%; Score 1654.8; DB 9; Length 159148;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1667; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 2045 CATTAAGCGCAAGCCTCAGAAAGACCTGGAAACAGATTCTGCCATCTCTGTGATCCCAAG 2104
Db 141963 CAGTAAGCGCAAGCCTCAGAAAGACCTGGAAACAGATTCTGCCATCTCTGTGATCCCAAG 142022
Qy 2105 ATGTCAGCCCTGCCCCAGCAATGCTGAATTTCTTCTACTTGGTCATCAAAAAGGAGT 2164
Db 142023 ATGTCAGCCCTGCCCCAGCAATGCTGAATTTCTTCTACTTGGTCATCAAAAAGGAGT 142082
Qy 2165 GTAATAGAAGTGAGGGAGCTGCTCCCTCATCTTCTTCTGAAGAAGAACCTTCTCTCT 2224
Db 142083 GTAATAGAAGTGAGGGAGCTGCTCCCTCATCTTCTTCTGAAGAAGAACCTTCTCTCT 142142
Qy 2225 TCTCTTCTCATGAATGGGCCCTTAGTGCCCTCAGAGATTGAGGACCCGACATCCCTC 2284
Db 142143 TCTCTTCTCATGAATGGGCCCTTAGTGCCCTCAGAGATTGAGGACCCGACATCCCTC 142202
Qy 2285 CACTCCAGAGTTGGTGTGACGAGTTTCAACTGCCCAACCTTGGCTCCACTATTGAA 2344
Db 142203 CACTCCAGAGTTGGTGTGACGAGTTTCAACTGCCCAACCTTGGCTCCACTATTGAA 142262
Qy 2345 TTTTTCAGACCCCAATTTCTTCATGCTGGAATGGGATTGCTGGACTTGGCAGCTTCT 2404
Db 142263 TTTTTCAGACCCCAATTTCTTCATGCTGGAATGGGATTGCTGGACTTGGCAGCTTCT 142322
```

QY 2405 TTCCCTCGCTTTGACTAGAACCGGACTCTTAATTTCTCAGGACAGACTAGCTGGCA 2464
|||||
Db 142323 TTCCCTCGCTTTGACTAGAACCGGACTCTTAATTTCTCAGGACAGACTAGCTGGCA 142382
|||||
QY 2465 CATTATCCCTACCTTAGTCTCTCTGACTCCCTGGGAAGTAATCTCTCTGTAATCTCTG 2524
|||||
Db 142383 CATTATCCCTACCTTAGTCTCTCTGACTCCCTGGGAAGTAATCTCTCTGTAATCTCTG 142442
|||||
QY 2525 TAAAGGTTTTGGGGGATAGGGTGTTTAAACCACCTCCAGAGCTTTCTCTCTC-TTTTTTT 2583
|||||
Db 142443 TAAAGGTTTTGGGGGATAGGGTGTTTAAACCACCTCCAGAGCTTTCTCTCTCTCTTTT 142502
|||||
QY 2584 TTTTCTGAAAAAGGAAAAAGCACACAGCACAAATTTCAAGCCATTTTCAGATCAAGACT 2643
|||||
Db 142503 TTTTCTGAAAAAGGAAAAAGCACACAGCACAAATTTCAAGCCATTTTCAGATCAAGACT 142562
|||||
QY 2644 CCAGAGTGTTGACAAAGTCCCTATTCTGAGAGTCCCTCAGAGAGCCATGGTGTTTAT 2703
|||||
Db 142563 CCAGAGTGTTGACAAAGTCCCTATTCTGAGAGTCCCTCAGAGAGCCATGGTGTTTAT 142622
|||||
QY 2704 GAAGAGAAGTAGTAGTCTCTGCTCCAGAGAGAGCTCTCTTTAACTCCCTCTCTCTT 2763
|||||
Db 142623 GAAGAGAAGTAGTAGTCTCTGCTCCAGAGAGAGCTCTCTTTAACTCCCTCTCTCTT 142682
|||||
QY 2764 GATGAATTTCTTAAGGCTGAAGGAATGAAGAGTGGGACATGGGCTAATCTTTATCCCT 2823
|||||
Db 142683 GATGAATTTCTTAAGGCTGAAGGAATGAAGAGTGGGACATGGGCTAATCTTTATCCCT 142742
|||||
QY 2824 TTTGTTAAACAGAGGAGCCAGCCATGGGCTGGGAGATCATAGCCCTTCCAGGAGAAATCC 2883
|||||
Db 142743 TTTGTTAAACAGAGGAGCCAGCCATGGGCTGGGAGATCATAGCCCTTCCAGGAGAAATCC 142802
|||||
QY 2884 TGTTCACTCCAGGCTATAGTAATATTACTATTGCAATTTGCAATTTGCAATATTCTGGTGG 2943
|||||
Db 142803 TGTTCACTCCAGGCTATAGTAATATTACTATTGCAATTTGCAATTTGCAATATTCTGGTGG 142862
|||||
QY 2944 TTTTCTTAATGTGAAGACTTACCAATGAATTTTAGATCATCTCCAGAGAGATTTT 3003
|||||
Db 142863 TTTTCTTAATGTGAAGACTTACCAATGAATTTTAGATCATCTCCAGAGAGATTTT 142922
|||||
QY 3004 TTTGCTCTCTCATCTTTTCCACAGTGTTCCTCGTGTGAGAGCTAAGGTAAGAGGG 3063
|||||
Db 142923 TTTGCTCTCTCATCTTTTCCACAGTGTTCCTCGTGTGAGAGCTAAGGTAAGAGGG 142982
|||||
QY 3064 GACACTTCTGCTGTTTAAACAGACAGTCCATATCTGAGGAGCCAGCAATATTTCTTAA 3123
|||||
Db 142983 GACACTTCTGCTGTTTAAACAGACAGTCCATATCTGAGGAGCCAGCAATATTTCTTAA 143042
|||||
QY 3124 ACTCATGGGGAGACAGAGATTTTGCCTTGTGAGGTGATGCTGTGCTCATATGTCCTA 3183
|||||
Db 143043 ACTCATGGGGAGACAGAGATTTTGCCTTGTGAGGTGATGCTGTGCTCATATGTCCTA 143102
|||||
QY 3184 CCCCCTGTCTTCATGACAGGAGTTGGAATGGGGGCTACATATGCCCTCTCTCCCGG 3243
|||||
Db 143103 CCCCCTGTCTTCATGACAGGAGTTGGAATGGGGGCTACATATGCCCTCTCTCCCGG 143162
|||||
QY 3244 TCTACAAGAGTTGTGGTTTTCATCTGATCCCTCCACCTCTCTCAGGAGGAGAGGGGC 3303
|||||
Db 143163 TCTACAAGAGTTGTGGTTTTCATCTGATCCCTCCACCTCTCTCAGGAGGAGAGGGGC 143222
|||||
QY 3304 CTGGTATCTCAGGAGATTTGTAATTCCTGTCTATCCCTCTCTATCCACCCCTGCT 3363
|||||
Db 143223 CTGGTATCTCAGGAGATTTGTAATTCCTGTCTATCCCTCTCTATCCACCCCTGCT 143282
|||||
QY 3364 TGATAATATGTAGCCCATACCCCAATTAAGTCTATATTAGACACCCCGAGCCAGTTT 3423
|||||
Db 143283 TGATAATATGTAGCCCATACCCCAATTAAGTCTATATTAGACACCCCGAGCCAGTTT 143342
|||||
QY 3424 CTGGCTGCTGCTTCTGCTGCTATTTTACAGAGAGGAGAAATCTTGCTATTTT 3483
|||||
Db 143343 CTGGCTGCTGCTTCTGCTGCTATTTTACAGAGAGGAGAAATCTTGCTATTTT 143402
|||||
QY 3484 TTTTTCATAATTTACTATTATGATGATTTTAAAGTGTGTTTTTAAAGGACAGAGTCTCTGTTA 3543

|||||
Db 143403 TTTTTCATAATTTACTATTATGATGATTTAAAGTGTGTTTTATTAAAGCACAGAGTCTGTTA 143462
|||||
QY 3544 GGGGTGGGAGGAATATTTGGGAGGGCTGGGTCTAGGGAAGGAATGGGGAAGCAAC 3603
|||||
Db 143463 GGGGTGGGAGGAATATTTGGGAGGGCTGGGTCTAGGGAAGGAATGGGGAAGCAAC 143522
|||||
QY 3604 ATTTTATTAAAGTGTACTATTGCTCTACTCTGCTTGTATTGTTTCAGAAATGGCAATACAA 3663
|||||
Db 143523 ATTTTATTAAAGTGTACTATTGCTCTACTCTGCTTGTATTGTTTCAGAAATGGCAATACAA 143582
|||||
QY 3664 TATAAAGTGATATATGTTTAAATGTAATAAATCTTAAATGAGTATTATTA 3713
|||||
Db 143583 TATAAAGTGATATATGTTTAAATGTAATAAATCTTAAATGAGTATTATTA 143632
|||||

RESULT 13
AC009564/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC009564 184090 bp DNA linear HTG 24-AUG-2002
Homo sapiens chromosome 10 clone RP11-413M23 map 10, WORKING DRAFT
SEQUENCE, 36 unordered pieces.

AC009564
AC009564-5 GI:8072465
HTGS_PHASE1; HTGS_DRAFT.
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 184090)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 10, clone RP11-413M23
Unpublished

2 (bases 1 to 184090)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Mollia,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesfaye,S., Toruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission
Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 184090)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lehoczy,J., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liou,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE	JOURNAL	COMMENT
-------	---------	---------

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7637235.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 36 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1	1060: contig of 1060 bp in length
* 1061	1160: gap of 100 bp
* 1161	2375: contig of 1215 bp in length
* 2376	2475: gap of 100 bp
* 2476	3802: contig of 1327 bp in length
* 3803	3902: gap of 100 bp
* 3903	5236: contig of 1334 bp in length
* 5237	5336: gap of 100 bp
* 5337	7037: contig of 1701 bp in length
* 7038	7137: gap of 100 bp
* 7138	8383: contig of 1246 bp in length
* 8384	8483: gap of 100 bp
* 8484	9908: contig of 1425 bp in length
* 9909	10008: gap of 100 bp
* 10009	11137: contig of 1129 bp in length
* 11138	11237: gap of 100 bp
* 11238	13218: contig of 1981 bp in length
* 13219	13318: gap of 100 bp
* 13319	16067: contig of 2749 bp in length
* 16068	16167: gap of 100 bp
* 16168	18833: contig of 2666 bp in length
* 18834	18933: gap of 100 bp
* 18934	22378: contig of 3445 bp in length
* 22379	22478: gap of 100 bp
* 22479	24946: contig of 2468 bp in length
* 24947	25046: gap of 100 bp
* 25047	26306: contig of 1260 bp in length
* 26307	26406: gap of 100 bp
* 26407	29679: contig of 3273 bp in length
* 29680	29779: gap of 100 bp
* 29780	33599: contig of 3820 bp in length
* 33600	33699: gap of 100 bp
* 33700	36502: contig of 2803 bp in length
* 36503	36602: gap of 100 bp
* 36603	39914: contig of 3312 bp in length
* 39915	40014: gap of 100 bp
* 40015	42987: contig of 2973 bp in length
* 42988	43087: gap of 100 bp

43088	46873:	contig of 3786 bp in length
48974	46973:	gap of 100 bp
48974	50544:	contig of 371 bp in length
50545	50644:	gap of 100 bp
50645	53900:	contig of 3256 bp in length
53901	54000:	gap of 100 bp
54001	58793:	contig of 4793 bp in length
58794	58893:	gap of 100 bp
58894	64521:	contig of 528 bp in length
64522	64621:	gap of 100 bp
64622	69423:	contig of 4802 bp in length
69424	69523:	gap of 100 bp
69524	75858:	contig of 6335 bp in length
75859	75958:	gap of 100 bp
75959	82538:	contig of 6580 bp in length
82639	82638:	gap of 100 bp
82639	89983:	contig of 7344 bp in length
89983	90082:	gap of 100 bp
90083	96628:	contig of 6546 bp in length
96629	96728:	gap of 100 bp
96729	106893:	contig of 10164 bp in length
106893	106992:	gap of 100 bp
106993	118325:	contig of 1337 bp in length
118330	118429:	gap of 100 bp
118430	130262:	contig of 11833 bp in length
130363	130362:	gap of 100 bp
130363	141462:	contig of 11100 bp in length
141463	141562:	gap of 100 bp
141563	154056:	contig of 12494 bp in length
154157	154156:	gap of 100 bp
154157	164998:	contig of 10842 bp in length
164999	165098:	gap of 100 bp
165099	194090:	contig of 18992 bp in length.

FEATURES	Location/Qualifiers
source	1. 184090

```
misc_feature /note="assembly_fragment"
misc_feature 1161..2375
misc_feature /note="assembly_fragment"
misc_feature 2476..3802
misc_feature /note="assembly_fragment"
misc_feature 3903..5236
misc_feature /note="assembly_fragment"
misc_feature 5337..7037
misc_feature /note="assembly_fragment"
misc_feature 7138..8383
misc_feature /note="assembly_fragment"
misc_feature 8484..9908
misc_feature /note="assembly_fragment"
misc_feature 10009..11137
misc_feature /note="assembly_fragment"
misc_feature 11238..13218
misc_feature /note="assembly_fragment"
misc_feature 13319..16067
misc_feature /note="assembly_fragment"
misc_feature 16168..18933
misc_feature /note="assembly_fragment"
misc_feature 18934..22378
misc_feature /note="assembly_fragment"
misc_feature 22479..24946
misc_feature /note="assembly_fragment"
misc_feature 25047..26306
misc_feature /note="assembly_fragment
clone_end:T7
vector_side:left"
misc_feature 26407..29679
```


Query Match		43.6%; Score 1618.2; DB 2; Length 184090;
Best Local Similarity		92.9%; Pred. No. 0;
Matches 1769; Conservative		0; Mismatches 113; Indels 23; Gaps 6;
QY	105	CCAGCAGCCAGCTTAAGCTTACTCTCTGTGAAAGGGGAAAGTATCCCTGTGTGAAAGC 164
DB	118262	CAAACAGCAGAGTTAAAGTTTTTTTTTTTGGAAAAGGGAAAGTTCCCTCGGAAAAACC 118203
QY	165	GGTTAAACTTGTGAGGGGTGCGGAGCTGAGTTCTCTCCCC--ATGCCAGGCGAATGGT 222
DB	118202	AGTTAAATTTGTGAGGGGTGCAAGACGTAAGTTCTCCCCCATGCCAGGGGAAATGGT 118143
QY	223	GTGCGCTTGTAGCTGTCTCAGAGCGCGCTCGACGTGTCTGAGGAGGCGCCCGAGAGGGGCG 282
DB	118142	GTGGAATTTTAGTGTCTCAGCAGCGGGTTGATTTTGTGAGGAAAGCCCTCGAGGGGGG 118083
QY	283	GGGAGGTGGCCACAC--GAACGGCGGTTCTGTAAAGAGAGCGTTGGGAAGATTCGATTCC 339
DB	118082	GGGAGGGGCTCCCAAGAAAGGTGGTTTTTATATAAAAGACATTTGGGAAGATTCGATTCA 118023
QY	340	GAGAAGAGG-AAGAACCGGATTGAAGAGAG-CCAGGCCGCTGAGGGGAGGGGGCTGCT 397
DB	118022	GAGAAGAGGAAGAGCGGGATTAAAGAGAGCCAGGCCCTTGAGGGGAGGGGGCTGCC 117963
QY	398	AAGATGGGCTCGGCTCCTCCGGGCGCTGCTCTCGGTGCGTTTTTTCATCCTTTGATCCC 457
DB	117962	AAGATGGGCTCGGCTCCTCCGGGCGGTGTCTTCGGTGGTTTTTC-TCGGTTGATCCC 117904
QY	458	GGGTGCCCTCTGTAACCTTGTCTCAGCATCTGGAATCAAGAGACCCATGGCATCTGAG 517
DB	117903	GGGTGCCCTCTGTAACCTCTGCTCAGCATCTGGAATCAAGAGTCCCATGGCATCTGAG 117844
QY	518	GTGCTTATGCTCTGGCATGCCCATCAAGAAATAGGCCATAGAAGTGTGATTCTCTCA 577
DB	117843	GTGCTTATGCTCTGGCATGCCCATCAAGAAATAGGCCATAGAAGTGTGATTCTCTCA 117784
QY	578	GGAGAGCAACATATAAAAGCAACCTCATCAGCCCTTGAAGGTGCCATCCAGTTAGGC 637
DB	117783	GGAG-----GGACAACCTCATCAGCCCTTGAAGGTGCCATCCAGTTAGGC 117739
QY	638	ATTACCCAGACTGTGGGAGCCTGAGTACCAAAACAGAGCGGTGATGTCTCATGCAAGAT 697
DB	117738	ATTACCCAGACTGTGGGAGCCTGAGTACCAAAACAGAGCGGTGATGTCTCATGCAAGAT 117679
QY	698	TTCTAGCTGGTTCAGAGTATCTTCTTCCAGTGAAGGAGCAACCTGACCCCTGCTCAT 757
DB	117678	TTCCACATGTTGAGAGTATCTTCTTCCAGTGAAGGAGCAACCTGACCCCTGCTCAT 117619
QY	758	CATCAATGACTTTCGTTTCAAGACCTATGCACCTGTTGCCCTTCGCTACTTCCGGGAG 817
DB	117618	CATCAATGACTTTCGTTTCAAGACCTATGCACCTGTTGCCCTTCGCTACTTCTGGGAG 117559
QY	818	CTATTTGGTATCGGGCCGAGTACTTGTATTCCTCTGAGTGAGCGCTGATGAA 877
DB	117558	CTATTTGGTATCGGGCCGAGTACTTGTATTCCTCTGAGTGAGCGCTGATGAA 117499
QY	878	CTCTGAGCTCTGGAGCTAGTGTTCCTTATCTATGTCGAGCAACCATGAGTTCAAT 937
DB	117498	CTCTGAGCTCTGGAGCTAGTGTTCCTTATCTATGTCGAGCAACCATGAGTTCAAT 117439
QY	938	ATTAGACAGTCCAAACATAAGAGCGCGAATTTCTGCAGAGCTGCTTCCAGGATACTAC 997
DB	117438	GTTAAGACAGTCCGACATAAAGAGCGGAGTTTCTGCAGAGCTGCTTCCAGGATACTAC 117379
QY	998	ATGAACCTCAACGAGAACCTCGGACTTTCGCTTAAATTTCTATGGACTGTACTGTG 1057
DB	117378	ATGAACCTCAACGAGAACCTCGGACTTTCGCTTAAATTTCTATGGACTGTACTGTG 117319
QY	1058	CAGGAGGTGGCAAGAACTTCGGATTGTGGTGAACAACTTTTACCAAGATCGGTA 1117
DB	117318	CAGACAGGTGGCAAGAACTTCGGATTGTGGTGAACAACTTTTACCAAGATCGGTA 117259
QY	1118	AAATGCATATCAATATGACCTCAAGGCTCAACCTCAACAGCGGGGCTTCCAGAAA 1177
DB	117259	AAATGCATATCAATATGACCTCAAGGCTCAACCTCAACAGCGGGGCTTCCAGAAA 117199
QY	1178	GAGCGAGAGAGCCTCTTCCACATTTAAAGACCTAGACTTCTTACAGACATCCCTGAT 1237
DB	117198	GAGCGAGAGAGCCTCTTCCACATTTAAAGACCTAGACTTCTTACAGACATCCCTGAT 117139
QY	1238	GGTCTTTTTTTGGATGCTGACATGTACACGCTCTCTGTAAAGACCTTCGACGCTGATGT 1297
DB	117138	GGTCTTTTTTTGGATGCTGACATGTACACGCTCTCTGTAAAGACCTTCGACGCTGATGT 117079
QY	1298	TTGCTGCTGCAGAGCTTCAAGATAAATGGATTATAGCCTCTTGTATCAATATCAATATA 1357
DB	117078	TTGCTGCTGCAGAGCTTCAAGATAAATGGATTATAGCCTCTTGTATCAATATCAATATA 117019
QY	1358	GATCATGCACAAAGAGCCTTTAAGCAGTGAACACAGTACTCAGTTGATATCTCAAGA 1417
DB	117018	GATCATGCACAAAGAGCCTTTAAGCAGTGAACACAGTACTCAGTTGATATCTCAAGA 116959
QY	1418	CCGGCCCCCAAGGCTCTGTATTCACAGCCATGGAATCCATCCAGGAGAGGCTCGA 1477
DB	116958	CCAGCCCCCAAGGCTCTGTATTCACAGCCATGGAATCCATCCAGGAGAGGCTCGA 116899
QY	1478	CGGGTGGTACCAGTACTGATGACCATATGGGTGGCATCCCTGCCCGGAATAGTAAA 1537
DB	116898	CGGGTGGTACCAGTACTGATGACCATATGGGTGGCATCCCTGCCCGGAATAGTAAA 116839
QY	1538	GGGAAAGGCTCTGCTTTATTTGGCATCATTTACAGTCTTACAGTCTTACAGTTTGT 1597
DB	116838	GGGAAAGGCTCTGCTTTATTTGGCATCATTTACAGTCTTACAGTCTTACAGTTTGT 116779
QY	1598	AAGAATTTGAGCAGCTCTTGGAAAGCCCTGGTATACATGACGGAGACACTGTCTCAGTGCAT 1657
DB	116778	AAGAATTTGAGCAGCTCTTGGAAAGCCCTGGTATACATGACGGAGACACTGTCTCAGTGCAT 116719
QY	1658	CGCCAGGCTTCTACGCTGAACGCTTCCAGCGCTTATGTGCAACACAGTATTTAAGAAG 1717
DB	116718	CGCCAGGCTTCTATGCTGAATGGTTTCCAGCGCTTATGTGCAACACAGTATTTAAGAAG 116659
QY	1718	ATTCCCTTGAAGCCTTCTCTTCCAAAAAGTTTCGGCTCTGGCTCATCTTCTCTCGGCGA 1777
DB	116658	ATTCCCTTGAAGCCTTCTCTTCCAAAAAGTTTCGGCTCTGGCTCATCTTCTCTCGGCGA 116599
QY	1778	GCAGGCTCCAGTGGCAATCTCTGCTTACTTACAGCCATTCGGTCTCTGGGGAACACAAG 1837
DB	116598	GCAGGCTCCAGTGGCAATCTCTGCTTACTTACAGCCATTCGGTCTCTGGGGAACACAAG 116539
QY	1838	GCACAGTGCACAAAGGAGGAGGAGCCAGGCTTACCTTGGTCTGCTGATGTT 1897
DB	116538	GCACAGTGCACAAAGGAGGAGGAGCCAGGCTTACCTTGGTCTGCTGATGTT 116479
QY	1898	TTACCTCAGACTCCACCTTTGGAGGAATACAGTGGGGCTCGCTTATTCCTGACCCCACT 1957
DB	116478	TTACCTCAGACTCCACCTTTGGAGGAATACAGTGGGGCTCGCTTATTCCTGACCCCACT 116419
QY	1958	TTTCTACCTCTAGTGGAGAGACTTTGCAAAATGCTTAACATAAAGT 2002
DB	116418	TTTCTACCTCTAGTGGAGAGACTTTGCAAAATGCTTAACATAAAGT 116374
RESULT 14		
BC031774		
LOCUS		
DEFINITION		
Mus musculus, phosphatidylinositol-4-phosphate 5-kinase, type 1		
ACCESSION		
BC031774		
VERSION		
BC031774.1		
KEYWORDS		
MGC.		
SOURCE		
house mouse.		
ORGANISM		
Mus musculus		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus		
REFERENCE		
1 (bases 1 to 2553)		

AUTHORS
TITLE
JOURNAL

Strausberg, R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowls, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 31 Row: 1 Column: 14
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5679330.

FEATURES
source

1. .2553
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:25310 IMAGE:4503697"
/tissue_type="Eye, retina, mouse strain C57BL/6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
408. .2048
/codon_start=1
/product="phosphatidylinositol-4-phosphate 5-kinase, type
1 beta"

CDS

/protein_id="AAH31774.1"
/db_xref="GI:21619323"
/db_xref="LocusID:18720"
/translation="MASASSGPAAGFSSLDACAPAGTAASGIKRATVSEGPSASVM
PVKIGHRVSDSGEYTKTSSALKAQLQITHTVGSLSKPERDVLMDQFVVE
SIFPSEGSLTPAHNIDNFRFYAPVAPRYRELFIIRPDYDLYSLCEPLIELSN
SGAGSLFYVSDDEFTIKVQHEAFQKLLPGYMNQNPRLPLPKFGLYCVQ
AGKNIRIVVMNNLPRSVKMHKDYLGKSYKRRASQKEREKTLPLFKDLFDLP
DGLFDADMTSALCKLORDCLVQSFINDYSILMSIHNDHHAOREPTSDTOYSAD
TRPAQKALYSTAMESIOGEARRGTEVEDHMGIPARNKGERLLLYIGLIDILQ
SYRFVKLHSHKALVHDGTVSVHRPGFYAERFQPMCNVYFKKIPKSPFKFRS
GPSFRSRSGPSGNSCTSQLMASGEHRAQVTTKAEVDPVHLGRPDVLPQTPPLEEISE
GSPVGPSPFVVGQPLQILNLSTLEKLDVAESEFTH"

BASE COUNT
ORIGIN

611 a 651 c 669 g 622 t

Query Match 41.68; Score 1543.4; DB 10; Length 2553;

Best Local Similarity 78.74; Pred. No. 0;
Matches 2050; Conservative 0; Mismatches 471; Indels 84; Gaps 14;

Qy 1 ATTAACAGCGCGTGGTAGGAAGCAGGAGAGGGCGCTTCGCTCCCTTTGGG---ACTTT 57
Db 30 ATAAACAGGAGTGGCTGGGAGGTGAGGAGGGGGGCGCATTCGCTTTGGGGCACCCTC 89
Qy 58 TCATCGCTCGTTTTTTTTCAGATGTGGCTTGGCTGGCGGCAAGGTCGCCAGCAGCAGC 117
Db 90 AGGTGCGCTAGGTTTCTCTTATAGGTTTGTCTATCCGAGACGCGGACTCCCGGGGCGCG 149
Qy 118 TTAGCTTACTCTCTCTGTGAAGGGGAAAGTATCCCTGTGGAAAGCGGTTAAACTGTG 177
Db 150 CTCA-AGTGTTCCTGTCTGTGTAAGAGAGGTAGCTTTGGCGGAGAGAGTTCAAGTGTCT 208
Qy 178 GAGGGGTGGGAGCTGAGTCTCTCCCATGCGCAGCGAGTGTGCGCTTACGCTG 237

Db 209 AGAAGGTGCGAGACTGCAGCTCTCCCATGTCGGGGCAAGTGGTGGCGGTCTAGAGCTTG 268
Qy 238 TCCAGGAGCCGGCTCGAGCTGTCTGAGGAGAGCCCGGAGGGGGGAGGTGGGCCACACA 297
Db 269 TCTTGGAGCTTGAACAACGCGCTGTGA-----GAAGCCCGCC 304
Qy 298 GAACGCGGTTCTTAAGAGACGTTGGGAAGATTCGATTCCGAGAGAGGAGAACCGG 357
Db 305 GAACGCGGTTCTGTGAAGAGACGTTGGGAAGATTCGATTCCGAGAGAGGAGAACCGG 364
Qy 358 ATTGAAGAGAGAGCCAGCGCTGAGGGGAGGGGCTCTTAAGATGGGTGGCGCTCTCTC 417
Db 365 ATTGAAGAGAGAGCCAGCGCTGAGGGGAGGGGCTCCCAAGATGGGTGGCGCTCTC 424
Qy 418 CGGCGCGTCTTCGCTGCGTCTTTCATCCTTTGATCCCGCGGCTCTCTCTGACCTT 477
Db 425 AGGCGCAGCG---CGGCGCGGTTTTCATCCTTTCATCCTGCGGGCGCTCTCTGTTACC-- 479
Qy 478 GTCCTCAGCATCTGGAATCAAGAGACCCATGGCATCTCAGGTGCTTATGCTCTGGCAT 537
Db 480 -GCAGCAGCATCTGGAATCAAGAGAGCCACAGTATCTCAGGAGACCTTCGCGCTCTGTCT 538
Qy 538 GCCCATCAAGAAATAGGCCATAGAAGTGTGATTTCCTCAGGAGACACATATAAAAA 597
Db 539 GCCTGTTAAAAAATAGGCCATCGAAGTGTGATTTCCTCTGGAGAGACTACATACAGAA 598
Qy 598 GACAACTCATCAGCCTTTGAAGGTGCCATCCAGTTAGGCATTTACCCACACTGTGGGAG 657
Db 599 GACAACTCATCAGCCTTTGAAGGTGCCATCCAGTTAGGCATTTACCCACACTGTGGGAG 658
Qy 658 CTTGAGTACCAAAACAGAGCGTGTCTCTCATGCAAGATTTCTACGTTGGTGGAGAT 717
Db 659 CTTGAGTACCAAAACAGAGCGTGTCTCTCATGCAAGATTTCTACGTTGGTGGAGAT 718
Qy 718 CTTCTTTCCAGTGAAGGGAGCAACCTGACGCCCTCTCATCTCACTCAATAGCTTTTCGTTT 777
Db 719 CTTTTCCTCAGTGAAGGAGCAGCACTGACCGGCTCACCATTACATGATGATTTGAT 778
Qy 778 CAAGACCTATGCACCTGTTGCTTCCGCTACTTCCGGAGCTATTGGTATCCGCGCCGA 837
Db 779 CAAGACCTATGCACCTGTTGCTTCCGCTACTTCCAGAGAGCTCTTTGGCATCCGCGCTGA 838
Qy 838 TGATTACTGTATTCCTCTGCAGTGAGCGCTGATTGAACCTCTAGCTCTGGAGCTAG 897
Db 839 TGATTACTGTATTCCTCTGCAGTGAGCGCTGATTGAACCTCTCCAAATCTCGAGCTAG 898
Qy 898 TGGTTCCCTATCTATGTGTCCAGGAGCAGTTCATTTAAGAGAGTCCCAACATAA 957
Db 899 TGGTTCCCTCTCTATGTGTCCAGTGATGATTAATTCATTAAGACCGTCCAGCATAA 958
Qy 958 AGAGCGGAATTTCTGCAGAGAGCTGCTTCCAGGATFACATACACCTCAACAGAGCC 1017
Db 959 AGAAGCAGAAATTTCTGCAGAGAGTGTCTTCCAGGATFACATACATTAACCAAAACCC 1018
Qy 1018 TCGACATTTGCTGCTTAATTTCTATGAGCTGTACTGTGTGAGGAGCTGGGCAAGACAT 1077
Db 1019 TCGTACTTTGCTGCGCAAAATTTTATGGATTTGTGATGTGTGCAAGCAGCGCGCAAGAACAT 1078
Qy 1078 TCGGATTTGCTGTATGAACAATCTTTTACCAAGATCGGTAAAAATGCATATCAAAATGA 1137
Db 1079 ACGAATTTGCTGTATGAACAATCTTTGCTCGGTGAGTCAAAATGCATATGAATATGA 1138
Qy 1138 CCTCAAGGCTCAACCTACAAACGCGGGCTTCCAGAGAGAGCGAGAGAGCTCTTCCC 1197
Db 1139 CTTGAAGGTTCAACTTACAAGCGCGGCTTCTCAGAAAGAAACAGAGAGAACTCTCCC 1198
Qy 1198 CACATTTAAAGACCTAGACTTTCTTACAAGACATCCCTCATGGTCTTTTTTGGATGCTGA 1257
Db 1199 CACCTTTAAAGACCTGGACTTCTTACAAGATATCCCTGACGCGCTATTTTGGATGCTGA 1258
Qy 1258 CATGTACACGCTCTCTCTGAAGACCTCGACGCTGACCTGTTGGTGTGCGAGAGCTTCAA 1317
Db 1259 CATGTACAGTCTCTCTGTGTAAGACTCTTACAGCGTGACTGTTTGGTGTGCGAGAGCTTCAA 1318

Qy	1318	GATAATGCATTACAGCGCTCTTGATGTCAATTCATATATAGATCATGCACACGAGAGCC	1317
Db	1319	GATAATGGACTACAGCGCTGTGATGTCCATCCACACATGACCATATGCACAGAGGAGCC	1318
Qy	1378	CTTAAGCAGTGAACACAGTACTCTAGTTGATACTCGAAGACGGCGCCCCCAAAAGGCTCT	1437
Db	1379	CACAGCAATGACACACAGTACTCGGCTGCACAGCGCAGACCAGCGCCACAGAAAGGCGCT	1438
Qy	1438	GTATTCCACAGCCATGGAATCCATCCAGGAGAGGCTCGAGGGGTGTGTACCATGGAGAC	1497
Db	1439	CTATTCCACAGCTATGGAATCCATCCAGGGCGAGGCTCGAGGGGGGCGCATGTGGAGAC	1498
Qy	1498	TGATGACCATATGGGTGGCATCCCTGCCCGGAATAGTAAGAGGGAAAGGCTCTCTGCTTTA	1557
Db	1499	TGAGGACCCATGGGTGGCATCCCTGCCCGGAATACAAAGAGGAAAGGCTCTGCTTTA	1558
Qy	1558	TATTGCAATCATTGACATTTACAGTCTTACAGGTTTGTTAAGAAGTTGGAGCACTCTTG	1617
Db	1559	TATTGGCATATTGATATTCTTCAGTCTTACAGGTTTGTTAAGAAGTTGGAGCACTCTTG	1618
Qy	1618	GAAAGCCCTGGTACATGACGGAGACACTGTCTCAGTGCATCGCCACAGGCTTCTACGCTGA	1677
Db	1619	GAAAGCACTGGTCCATGATGGGATACGGTGTCACTGCATCGTCCAGCTTCTATGCTGA	1678
Qy	1678	ACGGTTCACAGCGCTTCATGTGCAACACAGTATTTAAGAAGATTCCCTTGAAGCCTTCTCC	1737
Db	1679	CGGGTTCACAGCGTTTCATGTGCAACACAGTGTTCAGAAGATTCCCTTGAAGCCTTCTCC	1738
Qy	1738	TTCCAAAAAGTTTCGGTCTGGCTCATCTTTCTCTCGGGAGCAGGCTCCAGTGGCAACTC	1797
Db	1739	TACCAAAAAAGTTTCGGTCTGGCGCGCTTTCTCTCGGGATCAGGCCCCAGCGCAACTC	1798
Qy	1798	CTGCATTACTTACCAGCCATCGGTCTCTGGGACACAAAGGCACAAGTGCACAAAGAGC	1857
Db	1799	CTGC----ACCTCCCAGCTGATGGCCCTCTGGGGAACACAGACACAAGTGCACCAAGGC	1855
Qy	1858	AGAAGTGGAGCCAGCGGTTACCTTGGTCGCTCTGATGTTTACTCAGACTCCACCTTT	1917
Db	1856	GGAAGTGGAGCCAGATGACACCTTGGCGCTCTGTGTTACTCTAGACTCCACCTTT	1915
Qy	1918	GGAGGAAATCAGTGAAGGCTCGCCTATTCCTGACCCAGTTTTCACCTCTAGTTGGAGA	1977
Db	1916	GGAGGAAATGAAGTGAAGGTTCACTTGTTCCTGGCCCCAGTTTCTCAGCTTAGTTGGACA	1975
Qy	1978	GACTTTGCAATGCTAACTACAAGTACAACCTTGGAAAGCTTGAAGTTCGACAGTCA	2037
Db	1976	ACCTTTGCAATACTAAATTGAGTTCAACCTTGGAAAGCTTGATGTTGCAGAGTCA	2035
Qy	2038	GTTCCACCATTAAGCGCAAGGCTCAGAAGACTTGGAAACAAGATTTCGCATCTCTGTGA	2097
Db	2036	GTTCCACCACTAGCAGTGAAGGCTCAGAAGACTTGAACAAGACTCTGCTGTGCTTTGA	2095
Qy	2098	TCCCAAGATGTGAGCGCTTGCCCGACGAATGCTCAATTTCTTCTACTTGGTTCATCAAAA	2157
Db	2096	TCCCAAGATGTGAGCGCTTGCCCGACGAATGCTGAA----TCGTCCTTGGTTCATCAACA	2152
Qy	2158	AAGGAGTGAATAGAAGTGAAGGAGCTGCTCTCTCCATCTCTTCTTGAAGAAGAACCTT	2217
Db	2153	GAGGAGAGAAATGGAGCTAGCAGGGCTCTGAC-----ACTGT-----	2190
Qy	2218	CTCTCCTTCTCTTCCATGAATGGCGCTTAGTGCCCTCAGAGTTGAGGACCGCAGCA	2277
Db	2191	CTCCCCCTCTCTC-----TGGGCAAGTAGTGCCCTTGACAGTTGAGGACAGANTCG	2239
Qy	2278	TCCCCCTCACTCCAGAGTTGGTGGTACGGATTTTCAACTGGGCAACCCCTTTCGCCCTCCAC	2337
Db	2240	TCCCCCACTCCAGAAAGTGGTGGCGTGAATTTTCCACTG--CCTGCCATGGCTCTAT	2297
Qy	2338	TATTCAATTTTTTTCAGACCCCAATCTTTCATGCTGGAATGGGATGCTGAGACTTGGCA	2397
Db	2298	ACCTGAATTTTTTCAAAACCCCAATCTTGTGCTG--GACATGGGTTGCTGGACT--GGT	2354

QY	2398	GCTTTCTTTCCCTCGCTTTTGACTAGGAACGGGACTCTTAATTTCCTCAGACAGACTA	2457
Db	2398		
Db	2395	GTTTTCCTCCCTTTTACCTTTCTCTGGAGCTGGACTCTTCAATTCCTCAGGACGGCTG	2414
QY	2458	GCTGGACATATATCCCTACCTTACCTTTCTCTGTGACTCCTGGAAGAAATACCTCTCTA	2517
Db	2415		
QY	2518	ATCTCTGTAAGGTTTGGGGGATPAAGGGTGTAAACACCTCCAGCTTCTTCTTCT	2577
Db	2469	ATTTGTATAAAGGTTTGGGGGAAAGTAAGGATCATCTTACTCCCAACCTTCTTTTCC	2528
QY	2578	TTTTTTTTTCTGAAAAAGGAAAAA	2602
Db	2529		
Db	2529	TAAAAA	AAAAA
RESULT 15			
AK098097		2567 bp	linear
LOCUS			
DEFINITION			
ACCSSON			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

FEATURES
SOURCE

```

; : .2507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TRACH2005666"
/tissue_type="trachea"
/clone_lib="TRACH2"
/note="cloning vector: pME18SFL3"
BASE COUNT      598 a      660 c      681 g      628 t
ORIGIN

Query Match          41.4%; Score 1537.2; DB 9; Length 2567;
Best Local Similarity 79.2%; Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 453; Indels 85; Gaps 15;

QY      1 ATTACAGGCCGCTGGTTAGGAAGCAGCGAGAAGGGCGGTTCGCTCTTTGGG---ACTTTT 57
|| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 59 ATAAACAGCAGTGGCTGGGAGCTGAGGAGGGGGGCAATTCGCTTTTGGGGCACCTC 118
Qy 58 TCATGCGCTCGTTTTTTTTCAGATGTGGCTGGGCGCAAGTCCCAAGACCCAGC 117
Db 119 AGGTGCGCTAGGTTCCTTTAGGTTTGTCTATCCGAGACGCGGACTCCCGGGGCCCGG 178
Qy 118 TTAAGCTTACTCTCTCTGTGAAGGGGAAAGPATCCCTGTGGAAGGGGTTAAACTTGTG 177
Db 179 CTCA-AGTGTTCCTGTCTGTGAAGAGGTAGCTTTGGCGGAGAGCAGTTCAAGTTGCT 237
Qy 178 GAGGGGTGGGGAGCTGAGTTCCTCCCAATGCCAGCGAATGTGTGGCTTGAGCTGG 237
Db 238 AGGAAGGTGGCAGAGCTGCGAGCTCCGCCATGCCCGGCAAGTGTGTGGCTCCTGAGCTG 297
Qy 238 TCCAGGAGCGGCTCGAGCTGTCTGAGGAGGCCCGGAGGGGGGGAGGTGGCCACACA 297
Db 298 TCCTGGAGCCTGAACAACCGCTGGA-----GAGCCCGCC 333
Qy 298 GAACCGGGTTCTGTAAGAGACCTTGGGAAGATTCGATCCGAGAAGAGGAAGAACCGG 357
Db 334 GAACCGGGTTCTGTAAGAGACCTGGGAAGATTCGATCCGAGAAGAGGAAGAACCGG 393
Qy 358 ATTGAAGAGAGCAGCCGCTGAGGGGAGGGGCTGCTAAGATGCGTGGCTCCTC 417
Db 394 ATTGAAGAGAGCAGCCGCTGAGGGGAGGGGCTGCCAAGATGCGTCCGCTCCTC 453
Qy 418 CGGCGCTGCTTCGCTGGTGGTTTTTTCATCCTTTGATCCCGGCTCCCTCCTGACCTT 477
Db 454 AGGCCCCAGCG---CGGCGCGGGTTTTTCATCCCTTTGATGCCGGGCGCTGTGTGTACC-- 508
Qy 478 GTCCCTCAGCATCTGGAATCAAGAGACCATGGCATCTGAGTGGCTTATGCTCTGGCAT 537
Db 509 -GCAGCAGCATCTGGAATCAAGAGACCATGATATCTGAGGAGCCTTCCGGCTCTGTCAT 567
Qy 538 GCCCATCAAGAAAATAGGCCATAGAAGTTGATTTCTCAGGAGAGACAACATAAAAA 597
Db 568 GCCTGTTAAAAAATAGGCCATCAAGTGTGATTTCTCTCGAGAGACTACATACAAAG 627
Qy 598 GACAACCTCATCAGCCTTGAAGGTGCCATCCAGTTAGGCAATACCACACTGTGGGAG 657
Db 628 GACAACCTCATCAGCCTTGAAGGTGCCATCCAGTTAGGCAATACCACACTGTGGGAG 687
Qy 658 CCTCAGTACCAAAACAGAGCGTGTATGCTCATCAAGATTTCTACGTGGTTGAGAGTAT 717
Db 688 CTTGAGCACCACCAACAGCGTGAGCTGCTGATCGAAGACTTCTACGTGGTGGAGCAT 747
Qy 718 CTTCTTTCCCAAGTGAAGGGAGCAACCTGACCCCTGCTCATCACTACAAATGACTTTT 777
Db 748 CTTTTTCCCAAGTGAAGGGAGCAACCTGACACCGCTCACCATTTACAATGACTTTT 807
Qy 778 CAAGACCTATGCACCTGTTCGCTTCCGCTACTTCCGGGAGGTATTTGGTATCCGGCCGA 837
Db 808 CAAGACCTATGCACCTGTTCGCTTCCGCTTACTTCCAGAGAGCTTTCGGCATCCGGCC 867
Qy 838 TGATTACTTGTATTCCTCTCAGTGAGCGCTGATTTGAATCTGTAGCTCTGGAGCTAG 897
Db 868 TGATTACTTGTATTCCTCTCAGTGAGCGCTGATTTGAATCTTCCAAATCTCGAGCTAG 927
Qy 898 TGGTTCCCTATTTCTATGTGCCAGCAGATGAGTTTCATTATTAAGACAGTCCCAACATA 957
Db 928 TGGTTCCCTCTTCTATGTGTCCAGTATGATGATTCATCATTTAAGACCGTCCAGCAT 987
Qy 958 AGAGCGGGAATTTCTGCAGAGCTGCTTCCAGGATCTACATGAACTCAACAGAGACCC 1017
Db 988 AGAGCGGGAATTTCTGCAGAGCTGCTTCCAGGATCTACATGAACTCTTAACAAAACCC 1047
Qy 1018 TCGGACTTTGCTGCTAAATTTCTATGGACTGTACTGTGTGCGAGCGAGTGGCAAGACAT 1077
Db 1048 TCGTACTTTGCTGCCAAATTTTATGATTTGATGTACTGTGTGCAAGCGCGGCAAGACAT 1107
Qy 1078 TCGGATTTGCTGATGAACATCTTTTACCAGATCGGTAAAAATGCAATCAATATGA 1137
Db 1108 ACGAATTTGCTGATGAACATCTCTTGGCTTCGCTCAGTCAAAATGCATATGAATATGA 1167

Qy 1138 CCTCAAGCGCTCAACCTTACAAGCGCGGGCTTCCAGAAAAGAGGAGAGAGCCCTCTCTCC 1197
Db 1168 CCTGAAGGGTTCAACTTACAAGCGCGGGCTTCTCAGAAAAGACGAGAGAAACTCTCCC 1227
Qy 1198 CACATTTAAAGACCTAGACTTTTACAAGACATCCTCATGTCTTTTTTTTGGATGCTGA 1257
Db 1228 CACCTTTAAAGACCTGGACTTCTTACAAGATATCCTCATCGGCCCTATTTTTTGGATGCTGA 1287
Qy 1258 CATGTACACGCTCTCTGTAAAGACCTCGACGCTGACTGTTTGTGCTGCGAGAGCTCAA 1317
Db 1288 CATGTACAGTCTCTGTGTAAGACTCTACAGCTGACTGTTTGTGCTGCGAGAGCTCAA 1347
Qy 1318 GATTAATGATTTACAGCCCTTGTGATGTCAATCCATTAATATAGTCAATGACAAAGAGACC 1377
Db 1348 GATAATGGACTTACAGCCCTTGTGATGTCCATCCACAATGAGACCATGACAGAGGAGCC 1407
Qy 1378 CTTAAGCAGTCAAAACACAGTACTCAGTTGATCTGGAAGACGGCCCAAAAGGCTCT 1437
Db 1408 CACGAGCAATGACACACAGTACTCGGCTGACACGGCAGACACGCCCCACAGAGGCCCT 1467
Qy 1438 GTATTCCACAGCCATGGAATCCATCCAGGAGAGGCTCGAGGGGTGTACCATGGAGAC 1497
Db 1468 CTATTCCACAGCTATGGAATCCATCCAGGCGAGGCTCGACGGGCGGCACTGTGGAGAC 1527
Qy 1498 TGATGACCATATGGGTGGCATCCCTGCCCGGAATAGTAAAGGGGAAAGGCTTCTCTTTA 1557
Db 1528 TGAGGACCATATGGGTGGCATCCCTGCCCGGAATAAACAAGGAGAAAGGCTCTCTCTTA 1587
Qy 1558 TATTGGCATCATTTGACATTTACAGTTCTACAGTTCTTAAAGTTTGGAGACCTCTTG 1617
Db 1588 TATTGGCATCATTTGATATTTCTCAGTCTCAGGTTTGTAAAGTTTGGAGACCTCTTG 1647
Qy 1618 GAAAGCCCTGTGATACAGGAGACACTGTCTCAGTGCATCGCCAGGCTTCTACGCTGA 1677
Db 1648 GAAAGCACTGTGTCATGATGGGATACGTTGTCTCAGTGCATCGTCCAGGCTTCTATGCTGA 1707
Qy 1678 ACGGTTCCAGCGCTTCATGTGCAACACAGTATTTAAGTAGATTCCCTTTGAAGCCCTTCC 1737
Db 1708 CGGTTCCAGCGCTTTCATGTGCAACACAGTGTTCAGAAAGATTTCCCTTTGAAGCCCTTCC 1767
Qy 1738 TTCAAAAAGTTTCGGTCTGCTCATCTTCTCTCGGGAGACGCTCCAGTGGCAACTC 1797
Db 1768 TACCAAAAAGTTTCGGTCTGCGCTGCTTCTCTCGGGATCAGGCCCCAGCGGCACTC 1827
Qy 1798 CTGCTATTACTTACCAGCATCGCTCTGTTGGGAAACACAGGACACAAGTGCACAAAGGC 1857
Db 1828 CTGC---ACCTCCAGCTGATGGCTCTGTTGGGAAACACAGGACACAAGTGCACCAAGC 1884
Qy 1858 AGAAGTGGAGCCAGCGCTTCACTTGTCTGCTGTGATGTTTACCTCAGACTCCACCTTT 1917
Db 1885 GGAAGTGGAGCCAGATGTACACCTTGGGCGCTCTGATGTCTTACCTCAGACTCCACCTTT 1944
Qy 1918 GGAGAAATCAGTAGGCTCGCTTATTTCTGACCCAGTTTCTCACCTCTAGTTGGAGA 1977
Db 1945 GGAGAAATAGTAGGCTTCCCTGTTCCGGCCAGTTTCTCACCTCTAGTTGGAGA 2004
Qy 1978 GACTTTGCAATGCTTAAGTACAACTTCAACCTTGGAAAGCTTGAAGTTGAGAGCTCAGA 2037
Db 2005 ACCTTTGCAATGCTTAAGTTCACCTTGAACCTTGGAAAGCTTGAAGTTGAGAGCTCAGA 2064
Qy 2038 GTTCACCCATTAAGCGCAAGCCCTCAGAAAGCTTGGAAACAGATTTCTGCCATCTCTGGA 2097
Db 2065 GTTCACCCATTAAGCGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 2124
Qy 2098 TCCCAAGATGTGAGCCCTTGGCAGCAATGCTGAATTTTCTTCTTCTTCTTCTTCTTCTTCT 2157
Db 2125 TCCCAAGATGTGAGCCCTTGGCAGCAATGCTGAATTTTCTTCTTCTTCTTCTTCTTCTTCT 2181
Qy 2158 AAGGATGTAAATAGAGTGGAGGAGCTGCTCCTCCATCTTCTTCTTCTTCTTCTTCTTCTTCT 2217
Db 2182 GAGGAGAAATGTGAGCTACGAGGCTCTGAC-----ACTGT 2219

